Validation of the Safety Behaviour Test (SBT):

Using an Interactive Tutorial to Remove Adverse Impact

A dissertation submitted in partial fulfilment
Of the requirements for
The Degree of
Master of Science
In
Applied Psychology
At the
University of Canterbury

By Christopher Watson
2019

Research Supervisors:
Associate Professor Christopher Burt, University of Canterbury
Dr Katharina Näswell, University of Canterbury

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thank you, Mum and Dad. You have been so incredibly supportive through my entire university career. I wouldn’t be where I am today without you both. I hope I’ve done you proud.

Thank you also to my partner, Steph. Steph, I’ve finished university! We can finally start our life together! You have been an absolute rock for me throughout my whole masters. No one else has had to endure even half the late-night dissertation ramblings that you have. I have no idea where I would be without you.

And finally, thank you to Chris Burt, my supervisor. Without your guidance, and tireless patience I wouldn’t have a dissertation.

Abstract

Workplace accidents occur at an alarmingly high rate in New Zealand. Workplace safety literature suggests that 90% of these accidents involve new employees (Burt, 2015). Currently existing safety behaviour assessments are prone to bias, and it is difficult to select safe employees during a job application process. Gamification of assessments appears promising as a solution to these biases. The Safety Behaviour Test (SBT) is an assessment developed for selection to take advantage of the benefits of gamification. The test places individuals in a virtual environment and measures their safety behaviour. Previous work validating the SBT (Burt, Crowe, & Thomas, 2018) found that video game experience had an adverse impact on the assessment. The SBT was only able to measure authentic safety behavior from individuals with computer game experience. Individuals with computer game experience also scored higher on the SBT than others. In order to remove the adverse impact, an interactive pre-assessment tutorial was created. A study of 60 individuals found that the new tutorial removed the difference in SBT score between those with computer game experience, and those without. A second validation study compared SBT score against acquaintance ratings of safety behavior. It is unclear whether a lack of computer game experience still has an adverse impact on SBT score, as no significant correlations between criterion variables and SBT scores were found. The need for measures of safety behavior and their practical use is discussed.

Introduction

Workplace Accidents

Every year an estimated 2.78 million deaths worldwide are attributed to work factors. In addition, there are between 272 million, and 475 million serious workplace accidents annually across the world (Hamalainen, Takala, & Kiat, 2017). Most of these deaths (both in quantity, and per capita) are from lower income countries. When these low- and middle-income countries are compared against high-income countries, it is apparent that Asia and Africa in particular, possess much higher rates of accidents and deaths (Hamalainen, Takala, & Kiat, 2017). These increased rates of death appear to be related to a surplus of labour, poor financial security, and a lack of employee rights (DeFreitas, & Marshall, 1998). The tragically higher rate of workplace death in poorer countries makes some sense considering these variables. However, if one examines the accident rate of high-income countries only, it becomes apparent that New Zealand has one of the highest rates of accidents and deaths of the high-income countries (Lilley, Samaranayaka, & Weiss, 2013).

Workplace accidents in New Zealand cost ACC $3.4 billion during the 2017/18 financial year (Accident Compensation Corporation, 2018). The cost to the economy is much higher than this, as ACC only covers 80% of wages, and the cost of medical treatment. The total financial cost of workplace accidents for businesses, and public services is estimated to be around 4% of GDP, or around $12 billion (Pezzulo, & Crook, 2006; International Labour Organisation, 2017). The cost to the taxpayer, and businesses is disheartening, but pales in comparison to the human cost in lives, health, and wellbeing. Individuals who must take time off work after suffering a workplace accident, are at risk of developing depression, which can reduce their ability to return to work, and begin a vicious cycle (MacEachen et al., 2010).

Lilley, Samaranayaka, and Weiss (2013) compared rates of occupational fatal injuries (from 2005 - 2008) between New Zealand, Australia, Canada, Finland, France, Norway, Spain, Sweden and the UK. New Zealand had the highest rate of work-related death out of the examined countries. The increased rate of accidents appears to be caused by factors other than the nature of New Zealand work. New Zealand's shocking rate of fatal workplace injuries was still the highest of the countries once the results were standardised to account for the large number of high-risk industries in New Zealand. If New Zealand’s high rates of workplace death are not due to the nature of the work itself, then there must be another factor involved. This factor appears to be one of attitude and compliance rather than of environment.

New Zealand Safety Culture

New Zealand’s economy has historically been based on agricultural exports. Until 1960 around a quarter of our GDP was derived from agriculture (Easton, 2016), and from 1950 to 1990 25-35% of New Zealand’ workforce was engaged in employment regarded as high risk (Briggs, 2003). However, New Zealanders attitudes towards health and safety did not reflect this increased level of risk. Prior to the Health and Safety in Employment Act 1992, health and safety law was haphazard, and number of separate acts existed, each specifically legislating different industries (Department of Labour, 2003). To enforce these laws, safety inspectors were given statutory powers, and common practice was to conduct “raids” on locations suspected to not be meeting their safety guidelines (Department of Labour, 2003). Health and safety inspectors would go into worksites unannounced and prevent non-compliant work from being carried out. Shutting off machines, closing unsafe areas, and forbidding work to be carried out until improvements had been made. Not surprisingly, health and safety was perceived as onerous, and employees tended to underestimate the level of risk they were exposed to (Gill & Shergill, 2004; Slappendel et al., 1993). During this time accident levels in New Zealand were much higher. Deaths occurred more frequently, and near misses were under reported (Evans, & Quigley, 2003; McNoe et al., 2005). The lackadaisical, ad hoc treatment of employee safety in New Zealand was costing lives.

In 1993 new legislation was passed to improve national safety standards. The Health and Safety in Employment Act 1992 provided broad rules for all industries to follow, enforcing the monitoring of hazards and accidents. The 1992 Act lead to an increase in reporting of accidents, but ultimately was not strict enough on businesses to create true lasting change (Evans, & Quigley, 2003). It was not until the Health and Safety at Work Act 2015 that businesses were responsible to actively identify and manage risks. Under the new 2015 Act both employees and businesses have an obligation to participate in workplace safety. In addition, punishments for non-compliance have been dramatically increased. Company directors can be found personally liable for safety breaches, and even significant prison time.

Despite the 2015 Act being in place for just a short while, it appears to have led to a decrease in the rate of non-fatal injuries over this time (NZ Statistics, 2018). While any improvement is heartening, there is still further work to be done in order to reduce New Zealand's work accident rate. In order to address the high accident rate, it is appropriate to examine the individual factors that can lead to an accident.

Models of Workplace Safety

Older models of work accidents place emphasis on the fault of the individual who had the accident. Heinrich's (1931) Domino Theory stated that an accident was similar to a row of dominoes where one domino toppling causes the next to tip, and so on. Each domino in the sequence represents a factor contributing to the accident, and as the last domino topples, injury occurs. Heinrich's theory is similar to more modern theories in that he considered that removing just one of the risk factors would prevent the fall of subsequent dominoes. However, Heinrich placed a disappointing emphasis on the "ancestry", and incompetence of the worker. Essentially attributing the first domino to the intellect (or lack of) of the worker.

Newer models of accident causation and safety behaviour recognise that accidents are the product of several connected factors. While 70 - 80% of accidents are consistently found to involve human error, this error is not often the cause of the accident, but rather a catalyst that sets off a number of preconditions (Kim, Na & Ha, 2011; Shappell & Wiegmann, 1996; Stanton & Salmon, 2009). Most modern models, such as the Systems-Theoretic Accident Model and Processes (STAMP), Human Factors Analysis and Classification System (HFACS) and Accimap (Leveson, 2004; Shappell & Wiegmann 2003, Reason 1990, Rasmussen, 1997) recognise this. These models are unquestionably useful, though each also possess unique caveats.

The Systems Theoretic Accident Modelling and Processes model (STAMP) (Leveson, 2004) is a systems-theory based accident model, and views accidents as the result of complex dynamic processes. Rather than events, STAMP models control systems and the constraints they place on components of the organisation. A STAMP model allows one to examine where it was in the organisational chain that control of component failures was not managed. Accidents occur when there is a failure in these components, their interaction, or an outside disturbance which is then not handled by the control system. Control systems have a variety of constraint types from O-rings in a fuel tank to managers in an office. The systems theory perspective views event-based models as fallible due to human error. When selecting an event in the chain, the individual constructing the model often has to exclude others. Leveson (2004) argues that this choice says more about the individual using the model than the accident itself. However, compared to the other models STAMP requires a much more in depth, and complete understanding of the system the accident took place within. In addition, while STAMP may be superior in the analysis of complex physical, and technical errors, it is less efficient at analysing human based systems.

Rassamens (1997) Risk Management Framework is used with ACCIMAP, another accident analysis model. Accimap is also systems based, and Rassamen describes accidents as waiting to be unleashed by a human error, rather than directly caused by a human. An employee acting one way may avoid an accident, but if the accident root cause is not attended to then the accident is still 'waiting'. Safety is considered a property of the system that stems from the interactions between individuals on a variety of levels, from government down to workers. Accimap analysis forms a tree of interacting events across the six distinct vertical layers that affect an organisation: Government, Regulators, Company, Management, Staff, and Work. The factors leading to an accident can be traced back through this tree to higher levels than simple human error. However, if all events are laid out in the tree, Accimap can become very complicated and convoluted. This reduces the worth of the system as a tool to create change, as more complex models will be less convincing, and poorly remembered when presented to the layperson (Hafer, Reynolds & Obertynski, 1996).

Reason's (1990) Swiss cheese model of accident causation suggests there is never one cause of an accident, but a number of overlapping "holes" in various layers of safety. An accident can only occur if all the holes line up, if even one of the layers of safety protects the individual then there is no accident. For example, if an individual runs down a ramp, slips over, and falls off an unguarded walkway edge then there has been a serious accident. However, in this hypothetical (and highly simplified) scenario had the individual walked more carefully, the slippery surface had been signposted, or a guard rail been put around the edge, then the accident wouldn't have occurred. It would take just one of these layers of protection to not have a corresponding gap in safety, in order to prevent the accident.

HFACS (Shappel, & Wiegmann, 2000) builds upon Reason's (1990) Swiss Cheese model by separating the layers of safety into 4 distinct taxonomies explaining how human error can end in accidents. These factors are: Organisational influences; Unsafe Supervision; Preconditions for unsafe acts; and Unsafe acts examines factors related to an accident across different levels, each level is a place where the accident could have been prevented through appropriate intervention/planning/protections. Each slice represents a different level of health and safety procedure within the organisation. Under this model, an accident is viewed as the result of multiple overlapping gaps in safety across the different 'slices'. Each of these holes are described as "resident pathogens" by Reason that only need a single opportunity to breach defences. An HFACS analysis begins with the accident and works backwards through each of these slices to determine where the 'hole' was. HFACS models are predicated on the idea that an accident does not occur without an unsafe act, without unsafe behavior.

New Employee Safety behaviour

As noted, human error plays a role in 70 - 80% of all workplace accidents (Kim, Na & Ha, 2011; Shappell & Wiegmann, 1996; Stanton & Salmon, 2009). While all the above models show that human error is just one part of an accident, they also show that human error is almost always necessary for an accident to occur. Given that human behaviour leads to the majority of accidents, it is worth exploring how accidents occur on a behavioural level, and what safety behaviours, and individual traits can reduce accident rates.

In Neal and Griffin's (2002) model of safety climate and safety behaviour, employee safety behaviours are categorised as either safety compliance, or safety participation. Safety compliance describes the individual's basic safety activities and procedures that keeps them safe, such as obeying safety signs, or by using appropriate safety equipment. Safety participation in contrast encompasses safety behaviours where an individual actively works to make their entire environment a safer place, rather than just enhance an individual's safety. Examples of safety participation are making suggestions to improve safety, and attending safety meetings.

Christiansen et al (2009) builds upon Neal and Griffins (2002) model and conducted a meta-analysis on the roles that person and situation factors play in an accident. The updated model includes a new category of interest: distal-person related factors. These are individual traits that can moderate performance of safety behaviours. Of particular interest are two variables: risk-taking, and safety attitudes. Risk-taking is defined as voluntary and conscious exposure to danger (Salminen, Klen, & Ojanen, 1999). As many as half of individuals who have a serious accident took a known risky action immediately before the accident (Salminen, 1994). Turner, McClure, and Pirozzo (2004) conducted a review of literature on the interaction of risk-taking tendency and injury. Overall, they found that risk taking behaviour correlates with increased chance of sustaining an injury.

Christiansen et al (2009) also identifies safety attitudes as an important factor in safety performance. Safety attitudes, or safety consciousness, describes an individual's awareness of the importance of safety (Barling, Loughlin, & Kelloway, 2002). Safety conscious individuals will be constantly aware of their own safety as they work, and act in ways to ensure their safety (de Koster, Stam, & Balk, 2011). Higher levels of individual safety consciousness correlate with an improved safety climate, which in turn correlates negatively with workplace accidents (Barling et al., 2002).

The factors of safety participation, safety compliance, safety consciousness and risk-taking appear to explain a large portion of the behavioural antecedents to accidents. However, these behaviours and constructs are not evenly distributed across employees, and higher levels of risk taking and lower levels of safety participation, safety compliance are often attributable to new employees. When the rates of accidents are broken down across employee tenure a very obvious trend emerges. New employees are implicated in 90% of accidents (Burt, 2015). The on-boarding of a new employee is often proceeded by a selection process. This HR process is an opportunity for the organization to screen job applicants in terms of their likely safety behaviour once employed. Arguably one of the most efficient ways of aiding businesses in reducing the level of workplace accidents is to ensure that they are selecting employees who will engage in their work with a high level of safety behaviour. Unfortunately, present safety measurement tools used for selection are limited in their effectiveness.

Current Safety Measurement

Currently, safety behaviour measurement tools are restricted to two assessment techniques. These tools will obtain either self-reported safety behaviour data, asking the applicant directly about their safety behaviours, and safety knowledge in a job interview, or obtaining a self-report of characteristics thought to correlate with safety behaviours, for example measures of compliance (e.g. Hogan Safe System) or risk avoidance (e.g Orion PreEmployment System PE3-SAFE). Table 1 provides a summary of available commercial measures of safety adapted from Barrett (2010).

Table 1.

List of Commercially Available Safety Measurement Tools

Commercial Product	Publisher	Contents
Employee Reliability Inventory	Bay State Psychological Associates Inc.	81 items - Self-report
Hogan Safe System	Hogan Assessment Systems Inc.	206 items - Self-report
Personnel Reaction Blank	IPAT Inc./Niche Consulting	92 items - Self-report
Onetest Work Safety Assessment	Onetest Pty Ltd.	Unknown
Orion Pre-Employment System PE3-Safe	Orion Systems Inc	Unknown
Situational Safety Awareness Test	Psyfactors Pty Ltd/ The Rogers Group/ Prospect Consulting	104 items - Self-report
Work Safety Assessment	Psych Press	Unknown
Risk Type Compass	Psychological Consultancy Ltd (PCL)	102 items - Self-report
Health and Safety Indicator 2009	Psytech International Ltd.	Unknown
Work Attitude Inventory	Psytech International Ltd.	Unknown
RMP Safety Inventory	RightPeople	Unknown
Dependability and Safety Instrument	SHL plc	Unknown
Safety Attitude Survey	Synergy Safety Systems	Unknown
Employee Safety Inventory	Vangent (Pearson) Inc. via Creative Organizational Design (Canada)	105 items- Unknown
Personnel Selection Inventory	Vangent (Pearson) Inc	64-144 items - Self-report

While an extensive discussion and description of each measure listed in Table 1 is not appropriate, as an illustration the Employee Reliability Inventory by Bay State Psychological Associates Inc. is an 81 item questionnaire assessing 7 attributes and taking 12-20 minutes to complete. The attributes include items covering safety behaviour and knowledge, and also 6 other safety adjacent correlates: Freedom from disrupted performance, courtesy, Emotional Maturity, Conscientiousness, Trustworthiness, and Long-Term Job Commitment. The measure produces a three-page report for each participant, and a graphic representing the individual's "Likelihood of Unreliable Behaviour". The axes of the graphic are without units, and no evidence is provided supporting the suggestion that the information contained within it is predictive of individual's actual behaviours.

Measures such as the Employee Reliability Inventory are problematic for two reasons. Firstly, self-report measures are prone to impression management and social-desirability (Arnold & Feldman, 1981; Feldman & Arnold, 1978; Van de Mortel, 2008). Impression management is an attempt by an individual to present an idealised version of themselves to others (Paulhus, 1984). Social desirability is the bias of individuals to answer in a way that conforms to societal norms. Individuals who are given a self-report measure are often prone to these biases, particularly if one or both of the following conditions are met: there is an obvious desirable option; and there is a benefit for responding in a specific way (Leary, & Kowalski, 1990; Villanova, & Bernardin, 1991). All of the commercial products reviewed by Barrett (2010) fulfil both of these criteria. If these measures are used as intended, as part of a selection process, then individuals may try to answer in a manner that they think will get them the job. Each of these measures also highlights an obviously desirable trait such as Trustworthiness (Employee Reliability Inventory), Irritable - Cheerful (Hogan Safe System), or Nonviolence (Vangent (Person) Inc.). In self-report measures affected by Social desirability or Impression management these biases can account for between 10 - 75% of variance in the obtained scores (Nederhof, 1985).

In addition to being prone to biases, the commercial measures of safety tend to not assess safety directly at all. Instead they typically only measure safety adjacent behaviours. For example, the Employee Reliability Inventory uses self-report to assess individuals on a number of personality scales for example Courtesy, Emotional Maturity, and Conscientiousness. These scales are then used to predict safety behaviours in the workplace. However, personality and safety knowledge correlates only weakly with workplace accidents (Christian et al., 2009)

It is clear to see then, that presently available psychometric safety assessments are prone to bias, and have poor criterion-related validity (CRV). CRV is essential as it is the gold standard for measures used in selection. A measure without CRV is not able to predict the construct of interest. Given the importance of safety for the general wellbeing of the workforce, having measures with good CRV is vital. Thus, alternatives to self-report measures need to be developed. The recent move towards developing gamified psychometric tool may provide an answer. Gamification is proposed as a potential solution to the present issues with safety behaviour psychometrics.

Gamification

Gamification describes the use of game elements in a non-game setting (Deterding et al. 2011). A broad variety of factors can be gamified such as learning, work, relationships, and most pertinent to the present study, assessments. There are many ways to apply game elements to assessments. Already existing assessments may be enhanced with gamified elements. These enhancements to assessments may be minor, such as a progress bar, animation, or sound. More elaborate enhancements may reframe the test entirely. Collmus and Landers (2015) added gamified narrative elements to a measure of conscientiousness, rewriting the test in terms of a first-person short story. Alternatively, an assessment can be crafted to take advantage of gamification from its inception. Such an assessment may use gamification to test an individual even as they learn (Kocadere & Caglar, 2015), or even place an individual inside a virtual world to obtain behavioural data from them.

Gamified assessments are relatively new to the gamification field, and those that are present have varied CRV results. While some gamified assessment tools have been found to lack CRV (Kim & Shute, 2015), others are a valid measure of the intended construct (Shute, Ventura & Kim, 2013; Shute et al., 2016). This suggests that it is the unique content of each gamified measure that determines CRV, rather than the type of measure. Gamified assessments pose a unique hurdle for researchers, as both the systems used to measure, and the game systems themselves need to be developed carefully to attain CRV. However, despite extra difficulties, gamified assessments offer unique advantages that mean they are worth the extra development effort. Gamifying assessments increases an individual's engagement with the tool (Suh, Wagner, & Liu, 2015), which leads to increased time and effort spent on the assessment (Bailey, Pritchard, & Kernohan, 2015). Gamified assessments may be particularly useful in the measurement of previously difficult to assess variables, such as health and safety behaviours.

Gamified assessments can be superior to self-report assessments in terms of social desirability and impression management bias. Instead of asking the individual for information about their behaviours, a gamified assessment can simply measure behaviour. For example, risk-taking is a behaviour very prone to social desirability bias, in a self-report questionnaire people will generally portray themselves more conservatively (Nederhoff, 1985). While some level of risk taking can be considered a positive trait for certain job roles, higher levels of risk-taking behaviours are associated with sensation seeking (Wagner, 2001), such as drug use, and with impulse control and violent behaviour (Zuckerman, & Kuhlman, 2000). Where self-reported risk taking would be prone to bias, a gamified assessment of risk taking such as the BART (Lejuez et al., 2002) is able to measure an individual's risk-taking characteristics more directly. The BART is a gamified computer assessment that presents individuals with 90 virtual balloons to inflate. Each time the participant "pumps" the balloon it expands slightly and increases in value by $0.05. Each balloon has an increasing chance to pop each time it is pumped, becoming worthless. At any time, individuals may trade their balloon for the amount their balloon is worth. The BART therefore will directly assess individual's risk-taking behaviours. Conservative individuals inflate their balloons fewer times than individuals with higher risk-taking. The number of balloon pumps in the BART correlates with adult sensation-seeking (Lejuez et al., 2002), adolescent risk-taking behaviours (Lejuez et al., 2003), psychopathy and impulsivity (Hunt et al., 2005). Gamified assessments, therefore, have the potential to directly tap into the behaviour they wish to measure, rather than indirectly like self-reports. This means that gamified assessments lend themselves incredibly well as a part of the job applicant selection process.

In addition to the reduction of bias, gamified assessments have several unique benefits. Gamified assessments need less training to implement than other solutions. There is no need to train raters or test givers as the test is entirely self-contained. Without the need for extensive training, the results of gamified assessments will also be free of rater bias (Clarke, 2009). In addition, the self-contained set up means that a gamified assessment can be given multiple times without investing in new resources, cutting down on material costs (Clarke, 2009).

Gamified assessments also appear to tap into authentic measures of behaviour more than pen and paper tests. The behaviour of individuals during a gamified assessment, particularly in a virtual world, is more reflective of their behaviour in life (Kozlov, & Johansen, 2010). This may be because during a self-report questionnaire an individual must recall how they would act in a certain situation, whereas in a virtual world, they just *do.* Eliciting job-relevant behaviour from applicants is helpful, as past behaviour is one of the best predictors of future behaviour (Ajzen, 1991).

Gamified assessments are also uniquely suited to assess safety behaviours because of their ability to be used as a work sample. Obtaining a work sample of safety behaviour would require placing an individual in an unsafe situation to assess their response. This is obviously an unethical, and an unacceptable way of obtaining a work sample of safety behaviour. A gamified assessment taking place in a virtual environment, without the risk of physical injury, allows potential employees to give a safety behaviour work sample in an ethical fashion.

The Safety Behaviour Test

In order to address the issues with present safety behaviour assessment, and to take advantage of the benefits of gamification, the Safety Behaviour Test (SBT) was developed at the University of Canterbury. The SBT is a gamified assessment tool designed to measure safety behaviour. The test uses a first-person viewpoint, and takes place in a virtual warehouse. The test-taker assumes the role of “fork-lift driver number 1” and is asked to retrieve goods using a forklift, and load them in to a container. The test features 35 different decision points, each recorded by the program. Safety and risk decisions made by individuals are used to create an SBT score, ranging from 0 - 13 (more information is provided in the method section under "*The SBT")*. Burt, Crowe, and Thomas, (2018) ran a validation study comparing the scores of 100 individuals on the SBT against ratings of their safety behaviours by acquaintances.

The scores of the participants on the SBT were normally distributed, indicating that the SBT had some merit as a measure. However, it appeared that videogame experience had an adverse impact on a person's SBT score. Participants who had played video games before had significantly higher SBT scores compared to those who had not played video games before (Burt, Crowe, & Thomas, 2018). In addition, CRV for the SBT was only evident when videogame experience was controlled. Table 2 presents the correlations between SBT scores and safety behaviours found by (Burt, Crowe, and Thomas (2018). The behaviour of video game players in the SBT seems to reflect their real-life behaviour more strongly than those who do not play video games.

Table 2.

Correlational Analysis between the SBT score measurement and Acquaintance safety measures (Burt, Crowe, & Thomas, 2018)

	SBT Score	
Safety Behaviour Measure	Acquaintances *n=100*	Controlling for months spent playing computer games *n=30*
Safety compliance	.13	.42*
Safety participation	.00	.22
Safety Voicing	.02	.25
Safety Consciousness	.08	.42*
Risk-Taking	-.08	-.41*
Rule-Bending	-.20*	-.46**

**p < .05, **p<.01*

Adverse Impact and the Safety Behaviour Test

As video game players have an advantage on the SBT over non-game players, the test is unsuited for use in a selection procedure with non-game players. The effect of an unrelated variable on employment procedures is known as “adverse impact” (Schmitt et al, 1997). If the SBT in its present format was to be used in a selection process, it would provide an advantage for video game players, who tend to be white (Embrick, Wright, & Lukacs, 2012). The present study sought to reduce the adverse impact of the SBT by making a new pre-assessment tutorial which aimed to teach non-game players the basics of the game playing skills required to complete the SBT, and thus make it suitable for use with individuals who do not play video games.

One point of improvement in the SBT that might level out the differences in individuals is the pre-test instructions. In the 2017 validation study of the SBT it used a series of text and picture instructions to describe how to use the test, followed by some more specific oral instructions. This may have left participants unused to using a video game system overwhelmed and under prepared to complete the test. Furthermore, an interactive tutorial is known within game design literature to be superior to a static tutorial (Pinelle, Wong, & Stach, 2008). Therefore, the creation of an interactive tutorial is likely to impart greater game control knowledge to individuals. Thus, the present research focuses on the creation and validation of a pre-assessment tutorial for the SBT to remove the adverse impact of a lack of computer gaming experience.

Present Research

The research involved two studies. In study 1 the new pre-assessment tutorial was tested and SBT useability data was compared to that obtained in 2017. The second study was a CRV study to determine if game player still had an advantage when completing the SBT.

Study 1

Method

Design

Study 1 compared usability data on the SBT version 2.0, with usability data obtained for SBT version 1.0. SBT version 2.0 has the new pre-assessment tutorial, and a number of within assessment modifications designed to overcome adverse impact found for SBT version 1.0 associated with computer game experience identified by Burt, Crowe, and Thomas (2018).   Individuals were tested on the SBT version 2.0 and then filled out the *SBT Usability Questionnaire* detailing their previous videogame experience, work experience, and usability ratings (See Appendix A). These data are compared with usability data collected by Burt, Crowe, and Thomas (2018) for SBT version 1.0. The current investigation is part of an ongoing study reviewed, and approved by the University of Canterbury Ethics committee, reference number HEC 2017/26.

Participants

Study 1 sourced participants using haphazard sampling. A variety of recruitment methods were used to obtain participants. Posters were displayed around the University of Canterbury calling for participants, participants were obtained through word-of-mouth, and through email, and online messages. Sixty participants were gathered using these methods: 22 from advertisements around the university, 20 through word of mouth, and 18 were solicited through online messages.

Demographic information

The demographics of participants in Study 1 are shown in Table 3.

Table 3.

Demographics of Study 1 participants

Variables	Participants *n=60*
Males	24
Females	36
Mean age	27.67
(SD)	10.61
Age Range	18 - 77

Materials

Materials used for Study 1 consisted of a computer, the SBT program (see below for more information about this tool), and a usability questionnaire (see Appendix A) asking participants demographic information, their computer game playing history, work history, and ratings of 5 different feedback scales on the usability of the SBT.

The Safety Behaviour Test

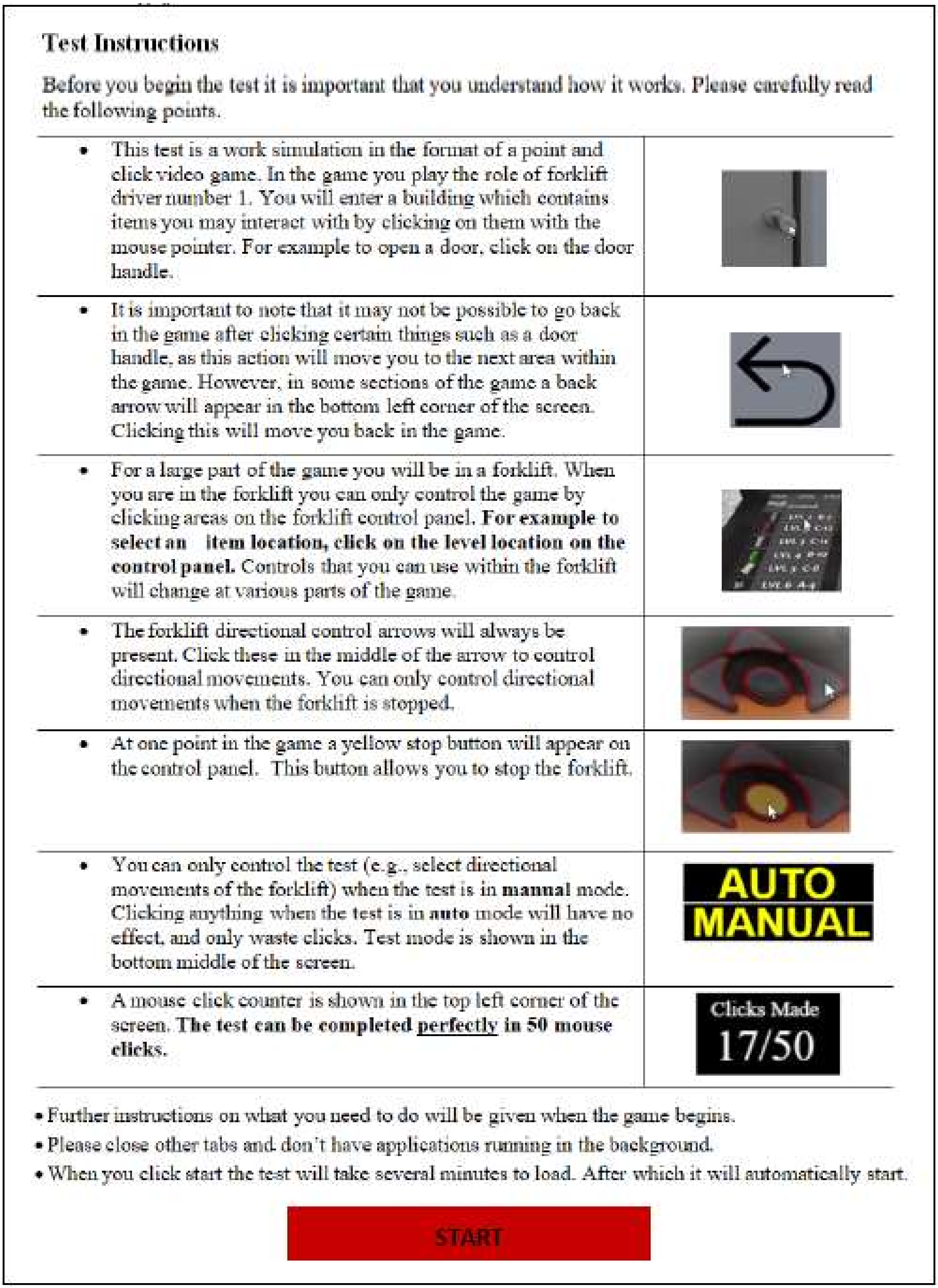
The SBT is a gamified assessment of safety behaviour. It is an animated three-dimensional point and click assessment, where the player must click on objects in the assessment to interact with them. In the assessment, players play through the eyes of “forklift driver number 1”, a forklift driver employed in a waste disposal warehouse. The player is asked to use a forklift to load items into a shipping container. Throughout the process of loading these items into the container, the player will come across 35 decision points some are simple assessment control decisions (e.g., opening a door), while 13 are about safety and the participant has the option to make either safe, or unsafe decision. Comparisons were made between versions 1.0 and 2.0 of the SBT. The differences between these versions are the pre-assessment tutorial, and the spoken instructions. Version 1.0 had a more rudimentary pre-assessment tutorial consisting of a single page of text and picture instructions which is shown in Figure 1, whereas Version 2.0 has an interactive pre-assessment tutorial, more detailed instructions, and other bug fixes and small changes (see Table 4 for a summary).

Table 4.

Summary of differences between SBT version 1.0 and Version 2.0

Area of difference	SBT Version 1.0	SBT Version 2.0
Pre-test Instructions	A page of text instructions with text on the side. See Figure 1.	Interactive tutorial consisting of five different pages, that each teach the participant a new aspect of the assessment. The tutorial covers: interacting with objects; driving the forklift; using the item display screen; stopping the forklift; and using the arrows in the corner of the screen to progress a scene. See Figures 2 - 7 for these scenes.
In-test Instructions	Assessment begins with the participant received broad instructions covering both their objectives, and how to control the forklift.	The assessment begins with the participant receiving instructions about their objectives, and then receiving more detailed instructions on controlling the forklift once they get into the forklift.
Feedback	No feedback during the assessment	Participant is thanked with an audio clip for pushing a warning button and thanked for their participation with a message following the end of the assessment.
Loading times	Longer load times	Faster loading times, optimised run more efficiently on slower computers.
Audio	Audio clips uneven	Audio clips equalised.
Buttons	Interactive clicking area of some buttons did not align with the button graphic.	Aligned the graphic of buttons with the area needed to click for interaction.
Stop button	To start the forklift after stopping, the stop button needed to be pushed again.	Either the stop, or the forward button can start the forklift after stopping.
Arrows to leave scenes	Only the check-in board, and the two safety signs had arrows participants could click to back out of the scene.	On top of the arrows at the three scenes with the signs, Arrows were added to scene at front desk so participant could proceed immediately after hearing the message.

As shown in Figure 1 below, the pre-assessment tutorial used in version 1.0 consisted of a single page of text and picture instructions. The static instruction page in version 1.0 explained to participants how they would be able to interact with the assessment. For example, the first instruction presented a picture of a door, and told participants it was possible to interact with objects in the assessment, such as the door, by clicking on them. Once a participant had read all the pre-assessment instructions, they clicked the red "START" button at the bottom of the page and began the assessment.



Test Instructions

Before you begin the test it is important that you understand how it works. Please carefully read the following points.

• This test is a work simulation in the format of a point and click video game. In the game you play the role of forklift driver number 1. You will enter a building which contains items you may interact with by clicking on them with the mouse pointer. For example to open a door, click on the door handle.	
• It is important to note that it may not be possible to go back in the game after clicking certain things such as a door handle, as this action will move you to the next area within the game. However, in some sections of the game a back arrow will appear in the bottom left corner of the screen. Clicking this will move you back in the game.	
• For a large part of the game you will be in a forklift. When you are in the forklift you can only control the game by clicking areas on the forklift control panel. **For example to select an item location, click on the level location on the control panel.** Controls that you can use within the forklift will change at various parts of the game.	
• The forklift directional control arrows will always be present. Click these in the middle of the arrow to control directional movements. You can only control directional movements when the forklift is stopped.	
• At one point in the game a yellow stop button will appear on the control panel. This button allows you to stop the forklift.	
• You can only control the test (e.g., select directional movements of the forklift) when the test is in **manual** mode. Clicking anything when the test is in **auto** mode will have no effect, and only waste clicks. Test mode is shown in the bottom middle of the screen.	AUTO MANUAL
• A mouse click counter is shown in the top left corner of the screen. **The test can be completed <u>perfectly</u> in 50 mouse clicks.**	Clicks Made 17/50

• Further instructions on what you need to do will be given when the game begins.

• Please close other tabs and don't have applications running in the background.

• When you click start the test will take several minutes to load. After which it will automatically start.

START

Figure 1. The Instruction Page for SBT Version 1.0.

SBT version 2.0 begins with an interactive tutorial that describes actions the participant will be able to engage in during the assessment, and provides an opportunity to practise aspects of the control of the assessment before actually starting the assessment. For example, the first tutorial screen presents the participant with a door, and tutorial text that tells the participant that they can click objects to interact with them (see Figure 2). This first tutorial screen invites the participant to click on the door handle in order to open the door. The participant is able to reset each tutorial slide, and try each action again. The participant is instructed to click an arrow to proceed to the next tutorial screen once they have practised sufficiently. See Figures 2- 7 below for the contents of the tutorial.

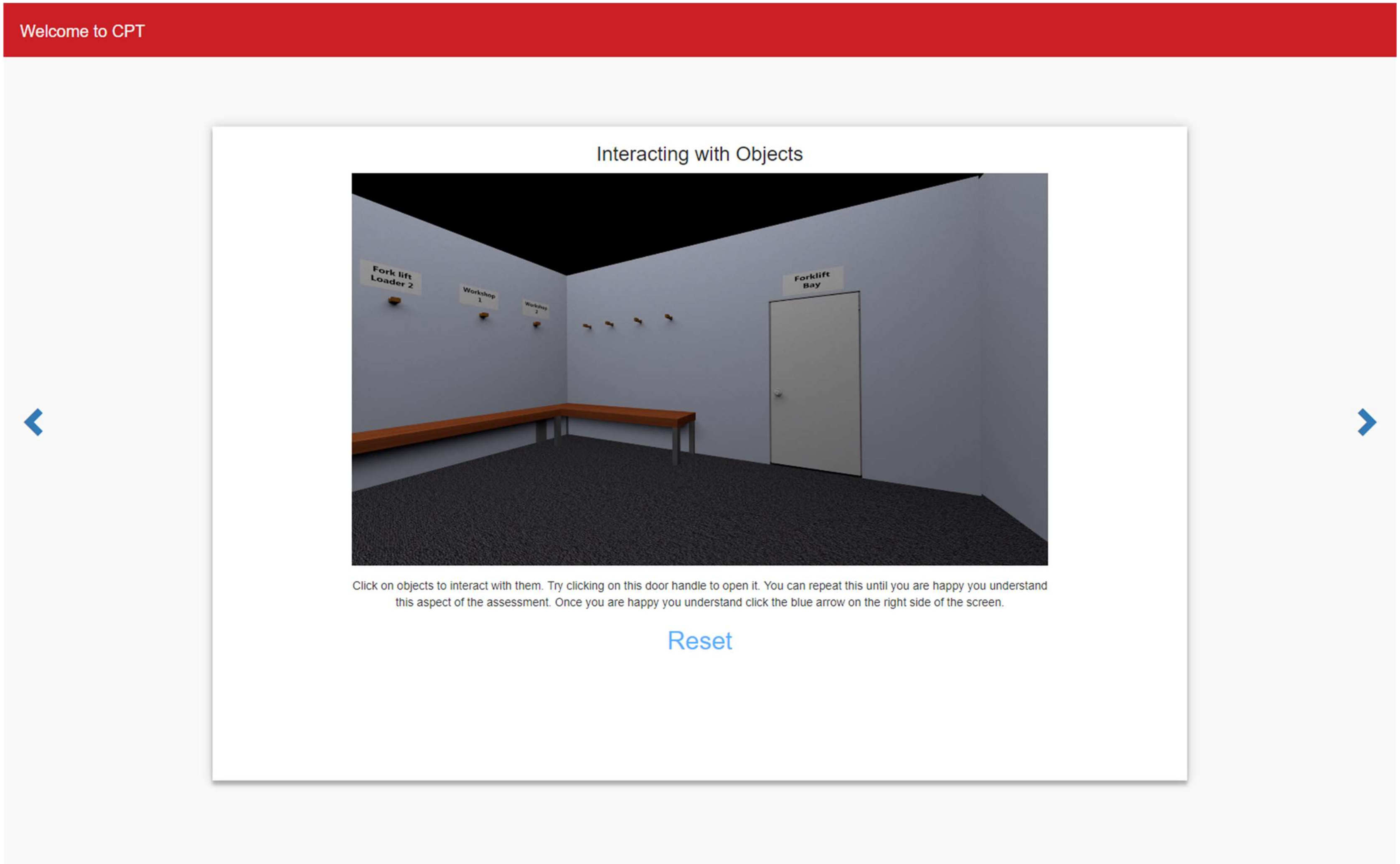


Figure 2. *SBT: First page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

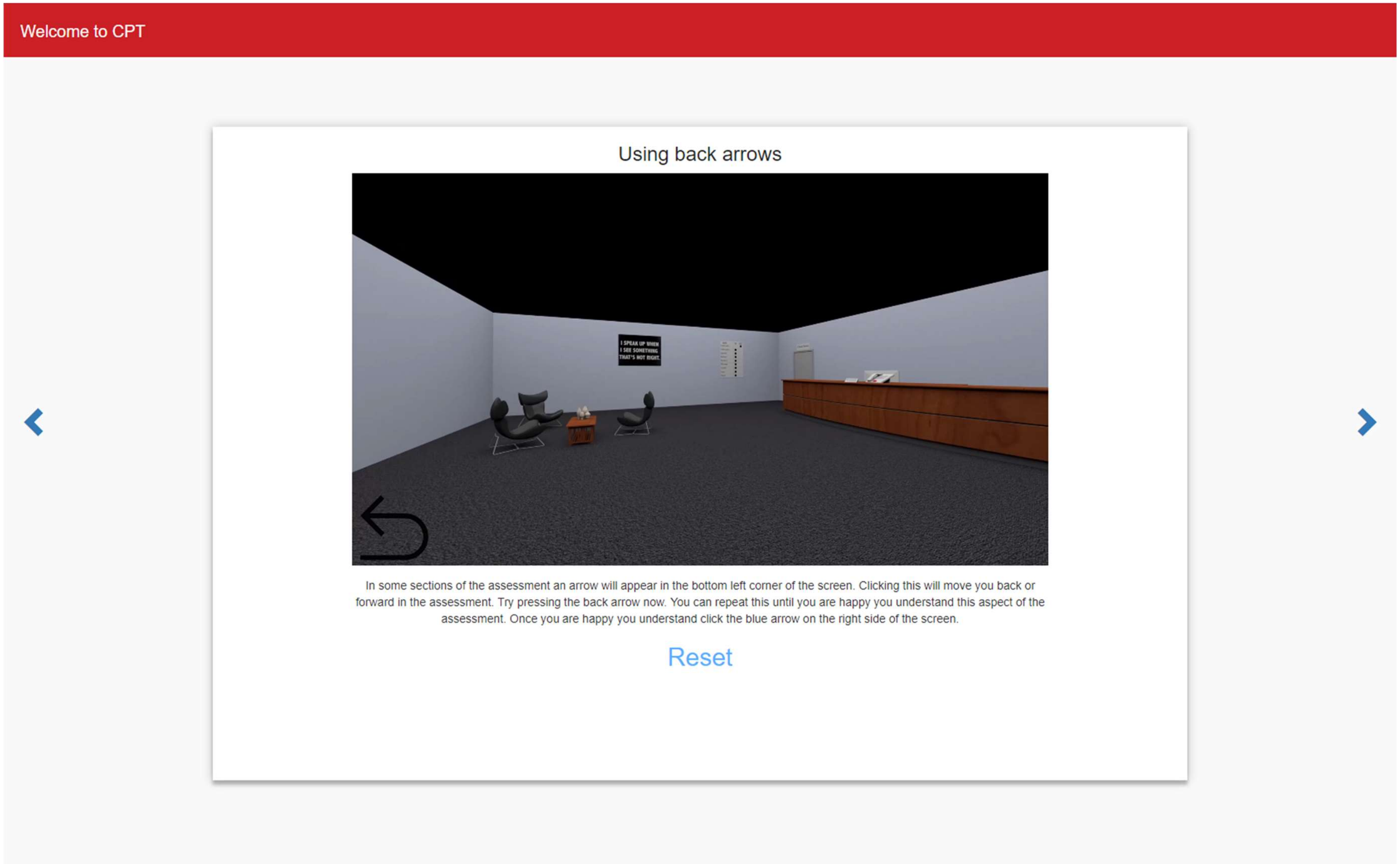


Figure 3. *SBT: Second page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

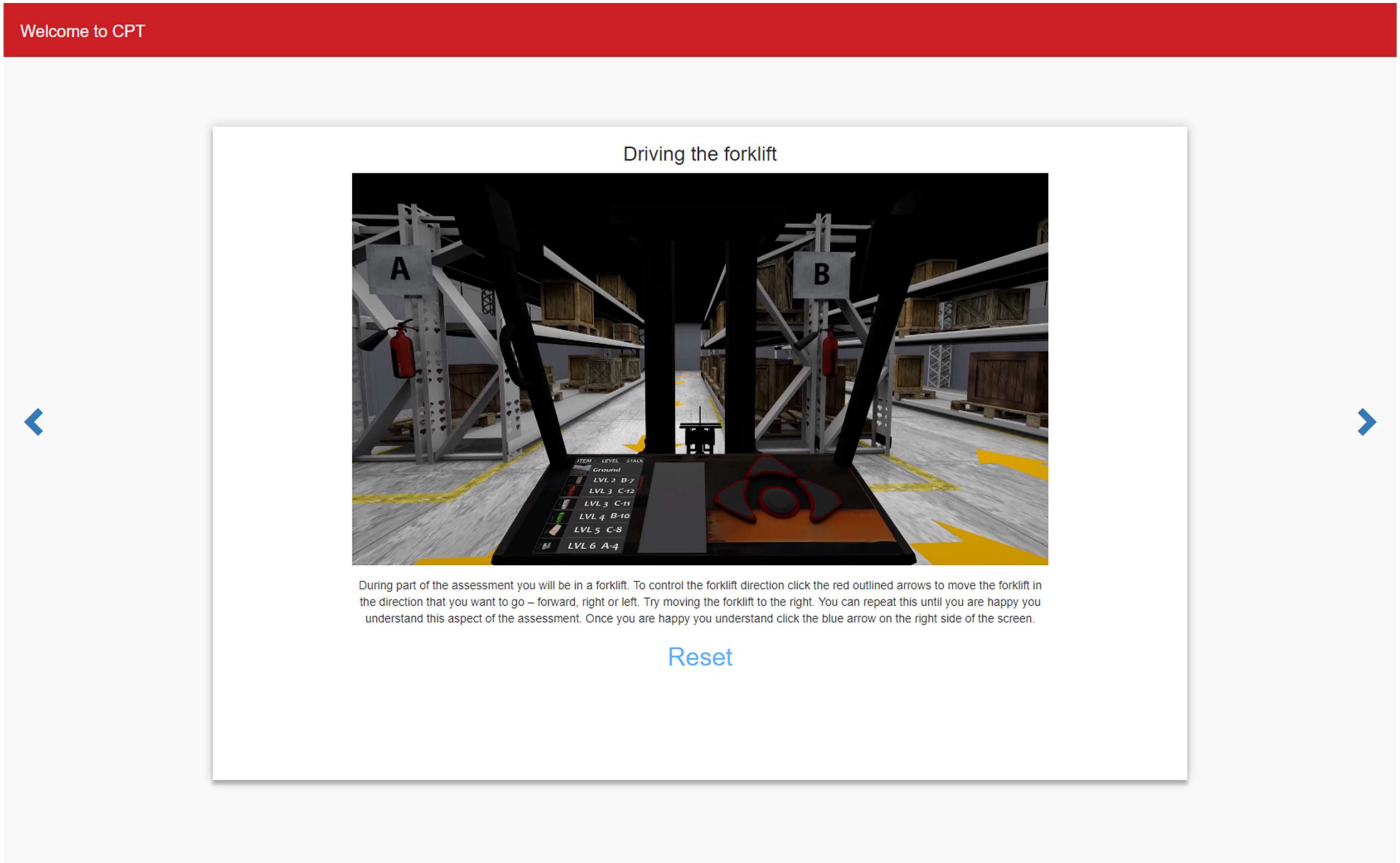


Figure 4. *SBT: Third page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

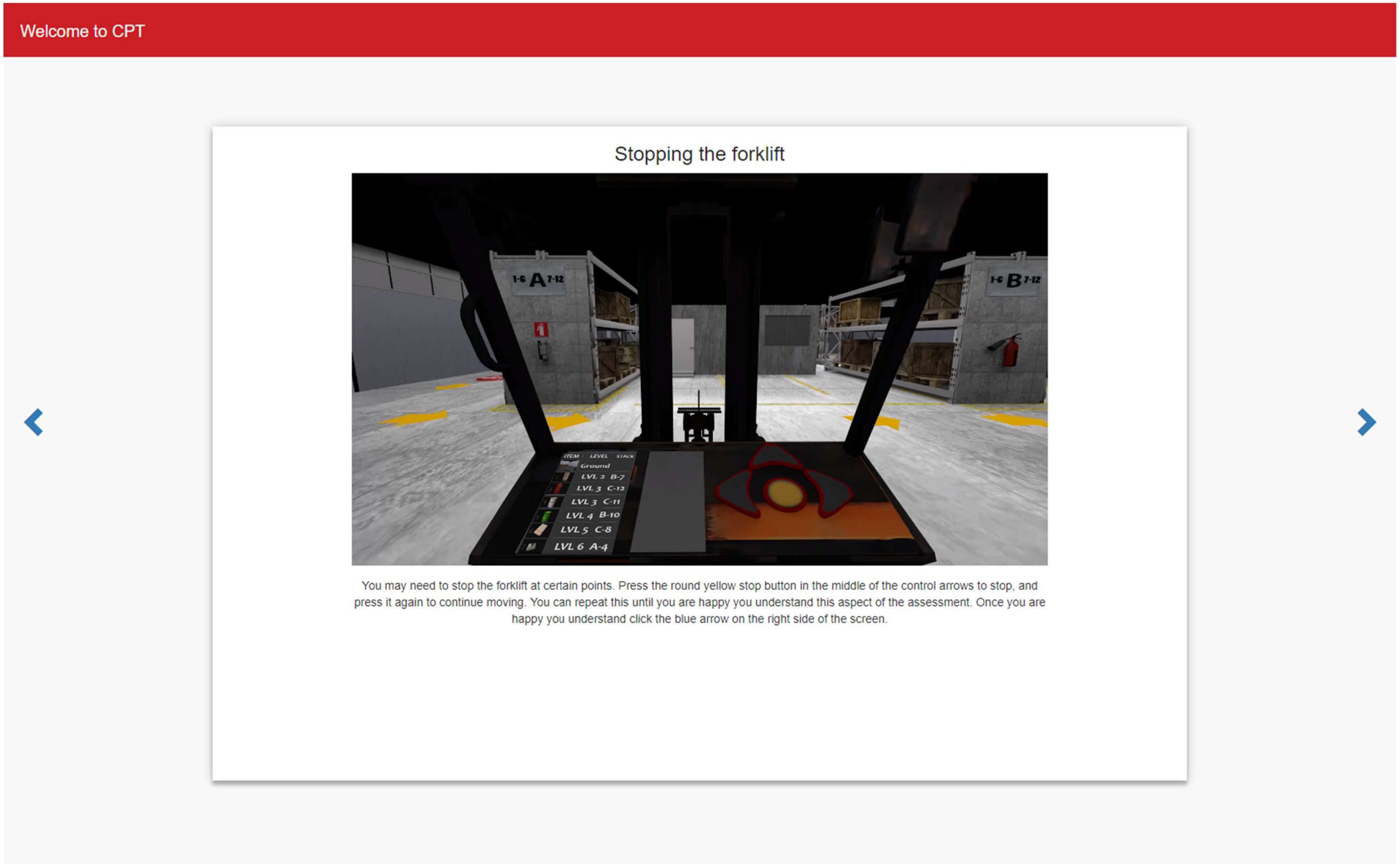


Figure 5. *SBT: Fourth page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

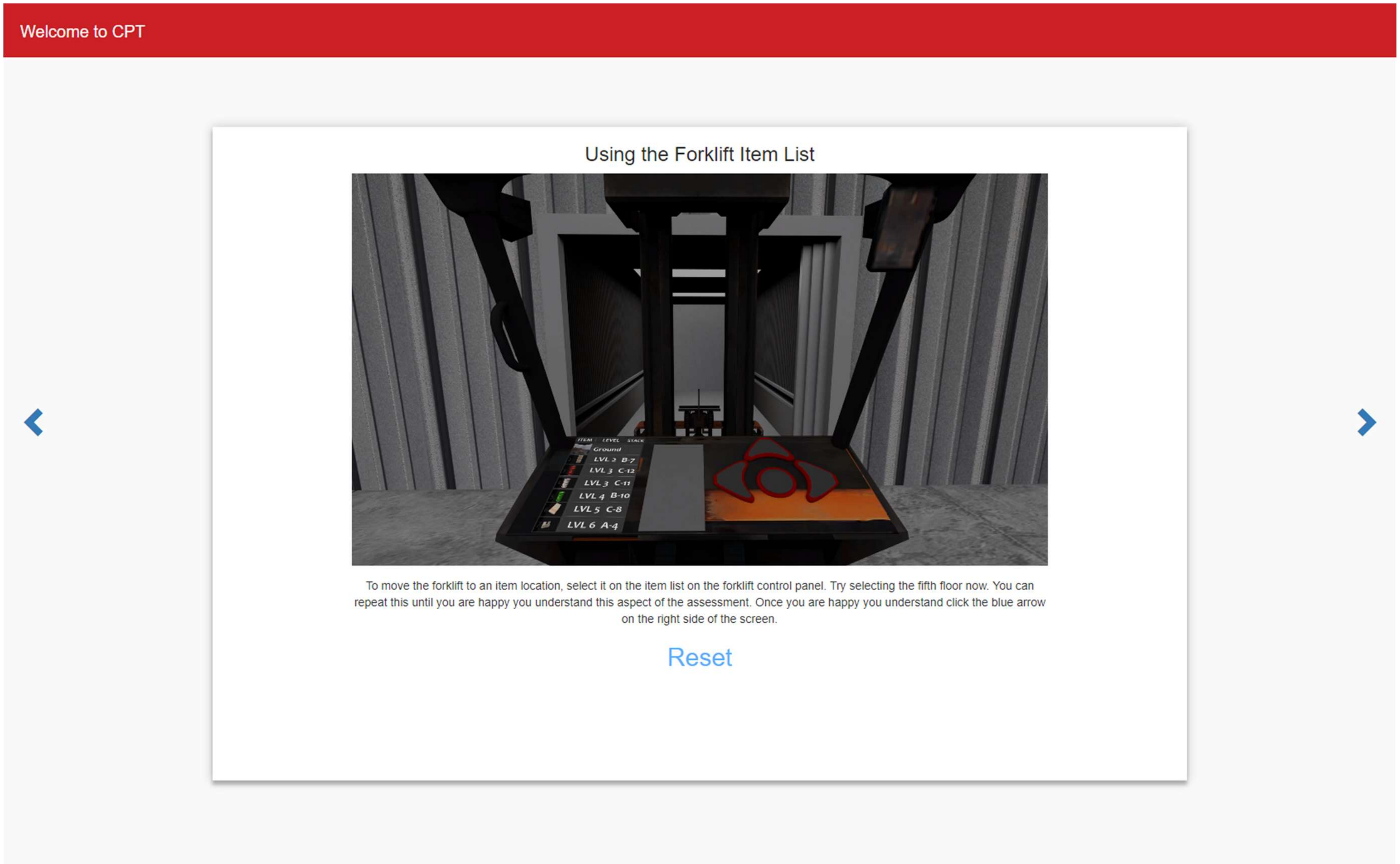


Figure 6. *SBT: Fifth page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

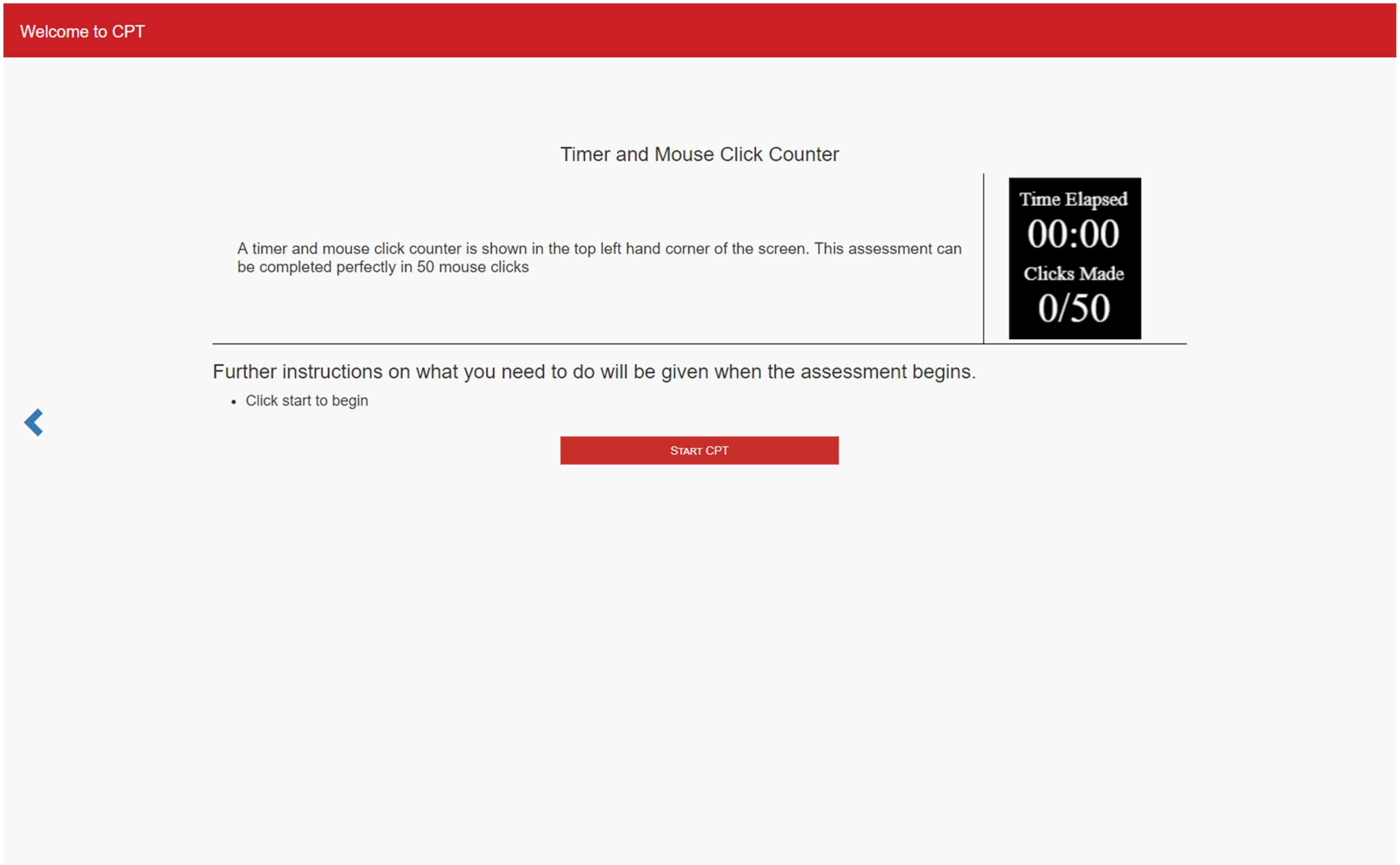


Figure 7. *SBT: Sixth page of Version 2.0 Interactive Tutorial (rotated 90 degrees).*

The other tutorial screens in version 2.0 allow the participant to practise aspects of the test: driving the forklift, using the item display screen, stopping the forklift, and using the arrows in the corner of the screen to progress a scene. Once the interactive portion of the tutorial has been completed, there is a final page (see Figure 7) stating that the number of mouse clicks will be recorded, and that the test can be completed in 50 mouse clicks. In addition, the final page tells the participant that they will be timed.

As the SBT is an assessment it is not possible, for assessment confidentiality reasons, to provide a complete description of it in this document which will become publicly available. However, selective screen shots are provided below to help the reader understand the assessment.

Once the assessment begins, the participant takes on the first-person view of an individual walking into an office, and up to a desk. They pause at a desk where they receive verbal instructions from a manager (see figure 8).

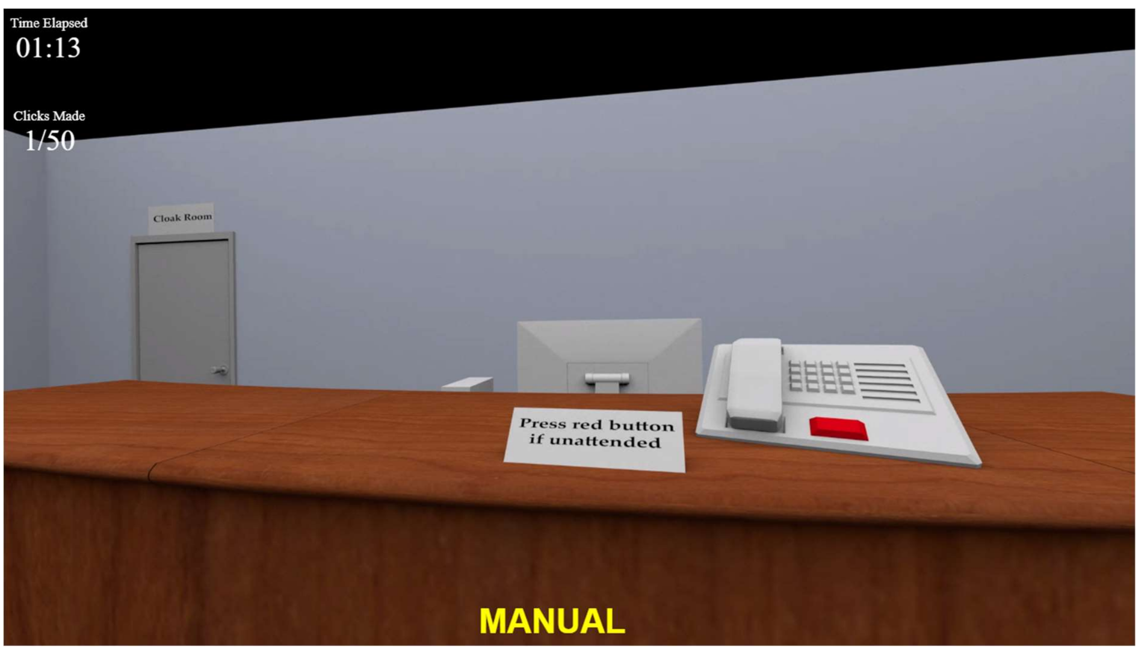


Figure 8. *Beginning of the SBT after walking up to the front desk, and clicking the Red Button.*

Once participants receive their instructions from, they continue past a sign-in board, through a cloak room full of safety gear and signs, and out to a forklift bay. While not in a forklift, participants can interact and move around the environment by clicking on the objects around them. For example, in Figure 9 the participant may click on any of the highlighted protective equipment, the safety instruction sign on the left, or click on the door to open it and continue through the warehouse. When the player is able to interact with the assessment and make decisions, the word "MANUAL" appears at the bottom of the screen (e.g. Figure 9). When the assessment carries out a decision selected by the player, the word "AUTO" is displayed instead (e.g. Figure 10), and the participant is unable to interact with the assessment until the action is complete.

Figure 9. *An Example of a Scene in the SBT with Interactive Components Highlighted for Emphasis.*

From the front desk, the participant moves through the building to the forklift bay. On the way to the bay they have a number of opportunities to demonstrate safety behaviours. Once in the bay, the participant selects a forklift, and the next part of the assessment begins.

After getting in a forklift, participants must navigate the warehouse, pick up items, and deposit them in a shipping container for dispatch. While the participant is in the forklift, they can only interact with the assessment by using the forklift controls (see Figure 10).

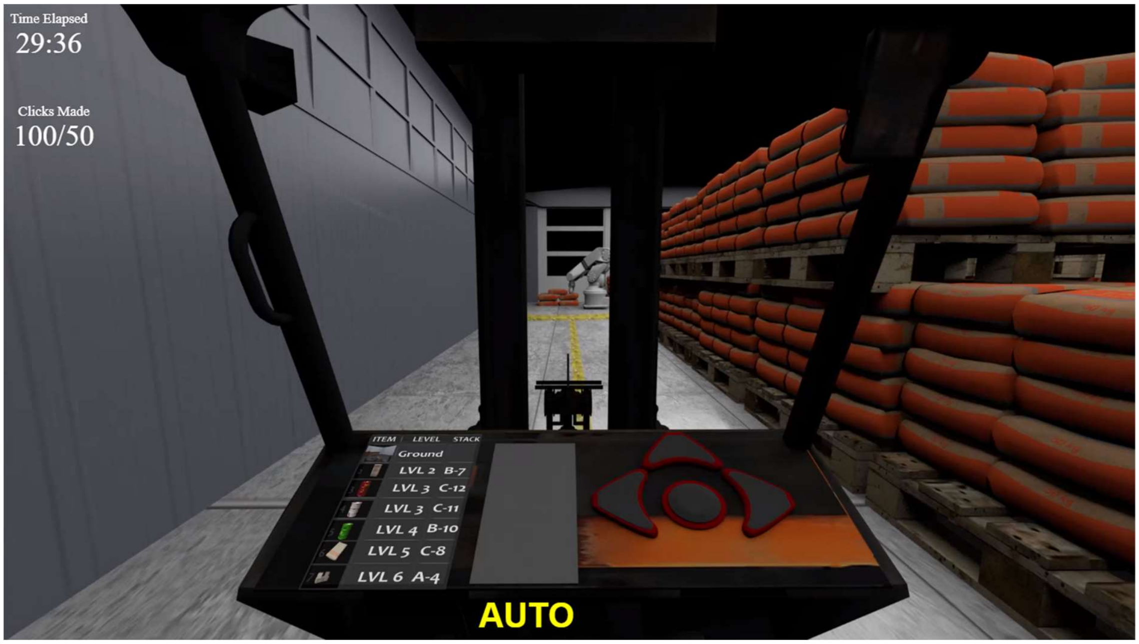


Figure 10. *An Example of a Forklift Scene in the SBT.*

The forklift controls consist of the item list used to select levels of the warehouse (on the left), the context sensitive button on the centre console (middle), the arrow keys to direct the forklift (on the right), or the stop button in the middle of the arrows. The context sensitive button changes depending on what level the forklift is on. Figure 11 shows what the centre button looks like on two different floors. On the fourth floor (top) it functions as a spill alarm, and in the forklift bay (bottom) it turns the lights on and off.

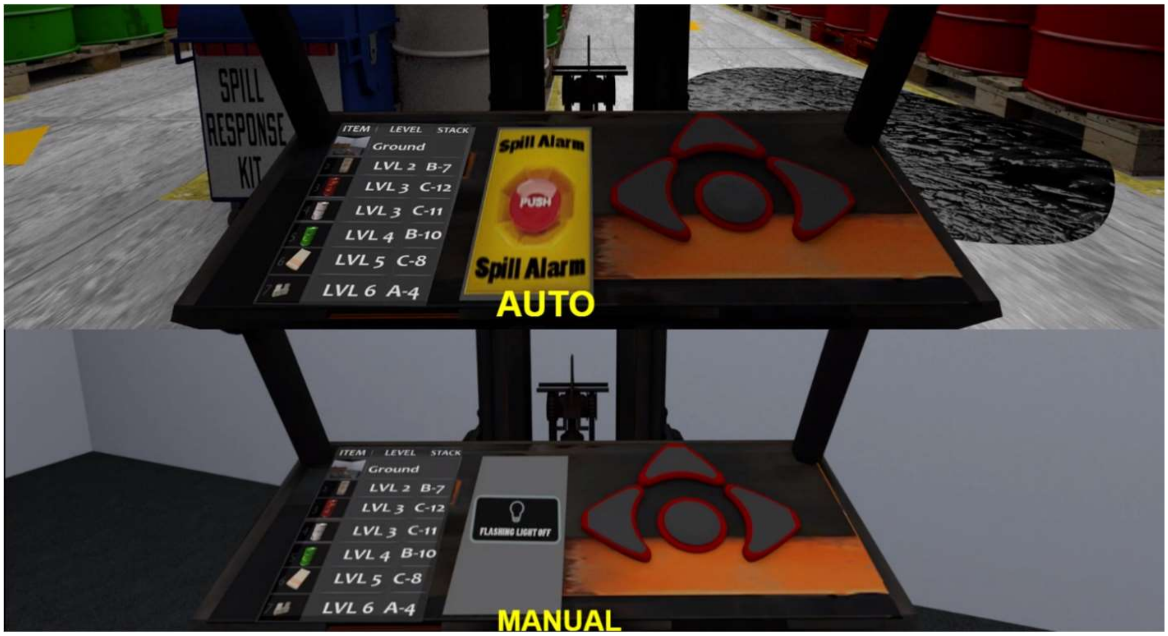


Figure 11. *Different uses of the context sensitive button. Top: On the fourth floor as a spill alarm. Bottom: In the forklift bay to turn on the flashing light.*

Throughout the SBT the participant has a number of opportunities to act both safely and unsafely. The participant is measured across 35 different decision points, such as control decision making, rule following, information seeking and reporting. 13 of these decision points concern the safety behaviours used to construct the final SBT score. Once the participants have loaded all five items into the shipping container they return to the front desk in the office. The unidentified manager asks: *Are you happy with the loading?* The participant is then given a final questionnaire (see Figure 12) asking them eight closed questions about safety in the warehouse.

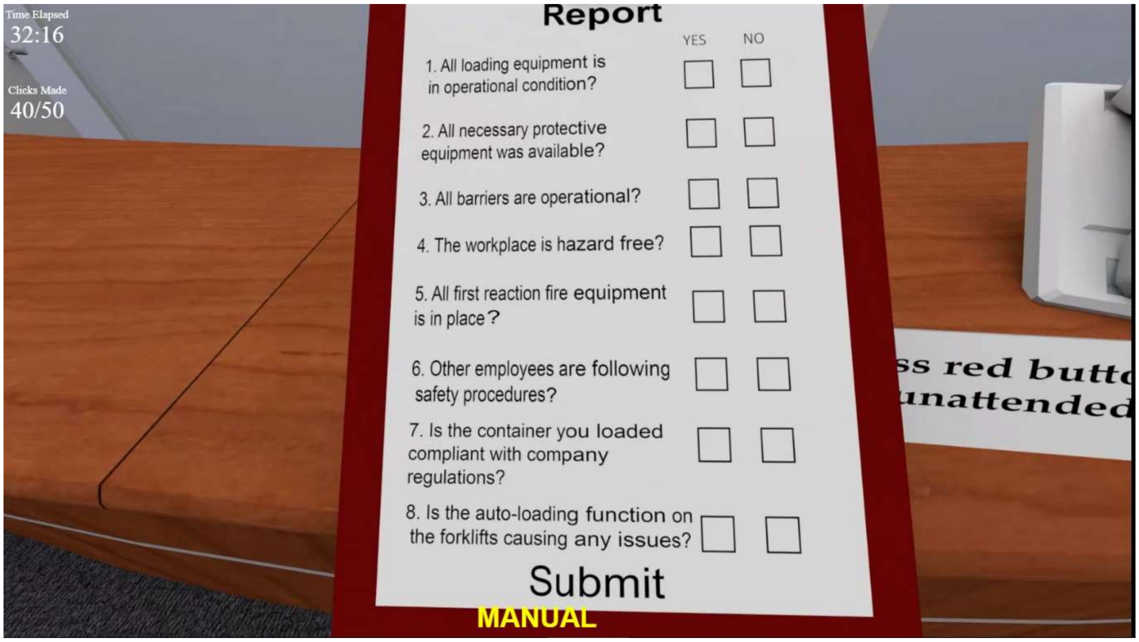


Figure 12. *Final Safety Questionnaire in SBT Version 1.0 and 2.0.*

SBT 1.0 and 2.0 differences in audio instructions.

As noted above in Table 4 each version of the SBT had a different set of instructions, each voiced by the same individual. The instructions for each version are shown below. Version 2.0 simplified the instructions slightly, and removed specific details about controlling the forklift.

Version 1.0: *"Hello forklift driver number 1. Sorry, I am up on level 6. It's good that you are here on time, there is only one job for you today. You have a shipment for disposal at the incinerator. The empty shipping container for the shipment is in loading dock C. A truck will take the loaded container to the incinerator as soon as you have finished loading it. I have already put the shipment items into the system, so when you get in a forklift the item list will be on the display screen. The new semi-automatic forklifts are working great, just click an item on the list and off you go to the relevant floor. Remember that control buttons appear when you need them. We have fixed the problem with the red right and left directional control arrows, and the central yellow stop button is working fine on all forklifts. Remember to load the items in the order shown on the list. The cloakroom is nice and tidy this week, so let's keep it that way. Don't muck around as the transportation firm will charge us if* *they have to wait, but be careful. When you have got the order loaded come back here and let me know. If you would like me to repeat the instructions, just click the red button again"*

Version 2.0: *'Hello forklift driver number 1. There is only one job for you today. I need you to put 5 items into shipping container C. I have already put the shipment items into the system, so when you get in a forklift the item list will be on the display screen. I'll give you a quick lesson on how to use the forklift once you get into it. We are busy today, so don't muck around, but be careful. Don't forget we are doing Health and Safety audit reports this month. If you would like me to repeat the instructions, just click the red button again'*

The SBT version 1.0 and 2.0 audio instructions diverge again once the participant gets into a forklift. Version 1.0 carries on to the next portion of the test without interruptions. Version 2.0 provides participants with further audio instructions from the same unidentified manager on how to use the forklift:

Hi driver, this semi-automatic forklift is controlled using the red arrow buttons, and the yellow stop button. You can only control the forklift when it is in 'Manual' mode. When you press an arrow the forklift will take you in that direction. Press the yellow stop button to stop, and press it again to continue moving. You can select which floor you wish to go to by selecting it on the item display screen on the left. Remember to load the items in the order that they appear on the list. Additional controls become available on the centre console at certain points.

The SBT Usability Questionnaire

The usability questionnaire was used to gather information about participants to examine the effect that individual differences had on assessment performance, and to gather the opinions of users on the usability of the assessment (see Appendix A) . The usability questionnaire has 5 areas for participants to fill out.

Demographic questions

Participants were asked their age and gender.

Computer game experience

Participants were asked questions to determine their level of experience with computer games, whether they had experience with point and click games similar to the SBT, and how often they played computer games.

Work experience.

Participants were asked about their present work situation, their present job tenure, their total job tenure, how many jobs they have had, and how many different organisations they have worked for.

SBT Usability

Participants were asked to rate the SBT on 5 usability factors. Each question was rated on a 7-point Likert scale. The questions were: "How understandable were the instructions given to you to use the SBT?" where 1 = "Not at all" and 7 = "Completely"; "How easy was it to control the forklift in the SBT?", where 1 = "Very Hard" and 7 = "Very Easy"; "How appropriate was the speed that the forklift moved in the SBT?", where 1 = "Very Inappropriate" and 7 = "Very Appropriate"; "Overall, how easy was it to complete the SBT?", where 1 = "Very Hard and 7 = "Very Easy"; and "How much did you enjoy completing the SBT", where 1 = "Not at all" and 7 = "Completely".

Comments

The participant was also asked whether they had any further comments, and a space was left for them to write.

Testing Procedure Study 1

Participants were given a page of information about the study, as well as a consent form (see Appendix A). Once their signed consent was obtained participants were instructed to complete the SBT with the instructions: *"The SBT is designed to be used as part of the selection process for jobs. Please try to put yourself in the mindset of a job applicant, and do as best as you can".* Participants would then begin the pre-assessment tutorial, practising each section as many times as they liked. After they felt ready, the participant began the assessment. After they had finished the SBT, participants completed the usability questionnaire (see Appendix A). Finally, after participants had completed both the SBT and the usability questionnaire they were given a $10 grocery voucher to thank them, and had the opportunity to ask questions.

Study 1 Results

Data Management

Data from each participants SBT results, and their feedback was entered into an SPSS file alongside data from Burt, Crowe, and Thomas (2018). Data was examined for outliers and missing data. No cases were missing any non-demographic data, of those cases that were missing data, none were missing over the 10% threshold to be removed. The data were examined to see if any cases had values more than 3 standard deviations away from the mean. If a value more than three standard deviations away from the mean was found, it was removed. The data was examined again, and the process repeated until all values fell within 3 standard deviations of the mean. Upon this examination of the data it became apparent that three participants had made many more mouse clicks during the SBT assessment than other participants (were outliers), and these cases were removed from the sample. See Table 5 for a summary of the outliers analysis.

Table 5.

Data Management of Study 1 Outliers

Highest case	*n*=	Sample Mean (SD)	# of clicks	Number of standard deviations away from mean.	Action
#11	63	107.71 (137.20)	1099	7.22	Case removed
#24	62	91.72 (52.57)	364	5.17	Case removed
#45	61	87.26 (33.9)	205	3.47	Case removed
#15	60	85.3 (36.63)	195	2.99	Finalised Sample

Pilot study feedback across videogame experience level

Difference in feedback between videogame players and non-videogame players were examined. One significant between group difference was found: that those with videogame experience found the assessment easier overall than those without videogame experience ($F(1,58) = 4.59$, p = .03). Table 6 shows the means and standards deviation for the variables examined.

Table 6.

Mean Feedback Compared across Levels of Videogame Experience

Feedback Measure	Mean (SD) Videogame experience *n= 40*	No videogame experience *n=20*	F-value
Understand Instructions	5.42 (1.15)	5.10 (1.41)	.91
Easy to control forklift	5.55 (1.41)	5.00 (1.62)	1.82
Forklift appropriate speed	5.37 (1.54)	4.70 (1.41)	2.67
Overall how easy	5.52 (1.03)	4.90 (1.11)	4.59*
Enjoyment	4.22 (1.76)	4.65 (1.13)	.96

* $p < 0.05$

SBT feedback across SBT version

2018 and 2019 feedback data on ratings of understanding instructions, easy of control of the forklift, appropriateness of forklift speed, how easy the assessment was to complete, and how enjoyable the assessment was were also compared to see if there was any significant differences between the SBT versions. See Table 7 for this comparison. There were significant differences between how much each version of the SBT was enjoyed. Participants enjoyed the newer version of the SBT significantly less than Burt, Crowe, and Thomas' (2018) version ($F(1,58) = 7.10, p < .01$).

Table 7.

Mean Feedback Compared across SBT Versions

Feedback Measure	Mean (SD) SBT 2.0 *n= 60*	SBT 1.0 *n=100*	F-value
Understand Instructions	5.31 (1.24)	5.06 (1.33)	1.43
Easy to control forklift	5.36 (1.49)	5.39 (1.500	.01
Appropriate speed	5.15 (1.52)	5.25 (1.44)	.19
Overall how easy	5.31 (1.09)	5.11 (1.49)	.84
Enjoyment	4.36 (1.58)	5.07 (1.63)	7.10**

** $p < 0.01$

SBT score compared across SBT version, and videogame experience.

Burt, Crowe, and Thomas, (2018) using the SBT version 1.0 found that individuals with video game experience had higher SBT scores than individuals without game experience ($F(1,98) = 5.12$, $p < .05$). The results from that research are presented alongside the present research in Table 8, where no significant difference between the mean SBT scores of videogame players and non-videogame players was found ($F(1,58) =. 62$, $p=.43$).

Table 8.

Mean SBT Score Compared across SBT Versions and levels of Videogame experience

	SBT score			
	Present study		Burt, Crowe, and Thomas (2018)	
	Video game experience	No video game experience	Video game experience	No video game experience
Sample size *n*=	40	20	30	70
Mean SBT score (SD)	8.40 (2.90)	7.80 (2.71)	8.91 (2.61)	7.66 (2.66)

* $p < 0 .05$

Summary

The Study 1 results suggests that the new SBT pre-assessment tutorial may have made the assessment more suitable for non-computer game players. Individuals with video game experience no longer show a score advantage on the SBT. The difference between versions of the SBT means that the new tutorial has improved the SBT user experience. However, participants in this study enjoy the new version significantly less than participants who competed Version 1.0. Of course, the two samples could be different on many unmeasured variables, and Study 1 did in fact sample many individuals who were working, and in contrast this study largely sampled from a student population. Study 2 will determine if the adverse impact found by Burt, Crowe, and Thomas (2018) has been removed.

Study 2

Method

Design

Study 2 examined the criterion related validity of SBT version 2.0, using participants from a variety of roles, across multiple organisations. Individuals were tested using the SBT version 2.0, and then given the *usability questionnaire used in Study 1* to fill out (see Appendix A). Participants were also given an *acquaintance questionnaire* which they gave to an acquaintance to answer questions about the participant's safety behaviours (see Appendix B). Participant SBT scores were correlated with their safety behaviour ratings from their acquaintance.

Participants

Study 2 sourced participants using haphazard sampling. Businesses in industries with a need for health and safety were approached and asked if they would allow their employees to take part in the study. Thirty-seven SBT participants, and thirty-seven acquaintances to rate the safety of the SBT participants were sourced from two different areas of a vineyard. Two cases were removed due to incomplete data.

The demographics of the thirty-five participants in Study 2 can be seen below in Table 9.

Table 9.

Demographics of Study 2 Participants

Variables	SBT participants *n=35*	AQ participants *n= 35*
Males	24	22
Females	11	13
Mean age	33.8	34.6
(SD)	13.7	12.0
Age Range	18-63	18-63

Materials

Materials used for Study 2 were similar to Study 1: A computer running the SBT program (see Study 1), a usability questionnaire (see Appendix A and Study 1) and an acquaintance questionnaire (see Appendix B).

The SBT

Version 2.0 of the SBT (See Study 1) was used in Study 2.

The Usability Questionnaire (see appendix A)

The same usability questionnaire was used in Study 2 as in Study 1.

The Acquaintance Questionnaire

The acquaintance questionnaire consisted of 5 sections (see Appendix B). The first section asked for information about the acquaintance, and their relationship with the participant.

General questions about you

Acquaintances were asked demographic questions, and about their relationship with the participant. In this section acquaintances were also asked to provide a general rating of the participant's safety risk on a 100-point scale, from "Not at all risky" to "Extremely Risky".

Criterion Variable Scales

The next four sections describe the scales to examine: *Safety Behaviour, Safety Consciousness, Rule Breaking, and Reliability.* Scales were responded to on a 5 point anchored Likert scale where 1 = "strongly disagree" and 5 = "strongly agree". In order to be used by acquaintances, these scales were altered from first-person language, to third person. Each case of an "I" pronoun was changed to "*…", where acquaintances were instructed that "*…" stood for the name of the participant who asked them to complete the questionnaire. For example, "I always take extra time to do things safely" became "*… always takes extra time to do things safely". Each of the four scales were presented in a random order, and the order of the questions within the scales were also randomised. Further details about section is provided below.

Safety Behaviour

Safety behaviour was measured using two 3-item scales, modified from Neal and Griffin (2006), measuring safety compliance and safety participation.

Safety compliance was defined as "the core activities that individuals need to carry out to maintain workplace safety" (Neal, & Griffin, 2006). An example item for safety compliance is *"*... always uses all the necessary safety equipment*". Whereas safety compliance focuses on personal safety, safety participation describes individual behaviours stemming from an active effort to improve the safety environment (Neal, & Griffin, 2006). An example item used for the safety participation scale is "*At work *... voluntarily carries out tasks or activities that help to improve workplace health and safety*". This scale was scored by adding all the ratings together and dividing the total by three.

The two subscales are reported to have good coefficient alphas. The safety compliance and safety participation measures were found by Neal and Griffin (2006) to have coefficient alphas of .93 and .89 respectively. In the present study the respective coefficients were .81 and .76.

Safety Consciousness

Safety Consciousness was measured using Westaby and Lee's (2003) 12-item Safety Consciousness and Risk-taking scale. Westaby and Lee (2003) defined safety consciousness as "*a positive attitude and awareness toward acting safely in general"*. Seven items on the scale concern safety consciousness. An example safety consciousness item from the scale is *"*... always avoid dangerous situations*". This scale was scored by adding all the ratings together and dividing the total by seven.Westaby and Lee (2003) reported a coefficient alpha of .83 for the safety consciousness measure, and the present study found a coefficient alpha of .82.

Risk-Taking

The second safety variable measured by Westaby and Lee's (2003) 12-item Safety Consciousness and Risk-taking scale is risk-taking, defined as "*an individual's willingness to engage in activities that knowingly have elements of physical danger*". Five items measured risk-taking. An example of one of the risk taking items is: "*In the past month *... has done some exciting things that* *other people might think are dangerous".* This scale was scored by adding all the ratings together and dividing the total by three. Westaby and Lee's (2003) reported a coefficient alpha of .77, and the present study found a coefficient alpha of .87.

Rule Bending

Chmiel's (2005) four-item rule bending scale was used to assess the participant's tendency to bend rules. For example, one of the items was: "*Work pressures means that *... sometimes bends the rules".* This scale was scored by adding all the ratings together and dividing the total by four. The rule bending alpha has been reported to have a coefficient alpha of .82 by Chmiel (2005), and the current study found a coefficient alpha of .81.

Reliability

Reliability was assessed using a developed seven-point scale, with 5 questions adapted from Cialdini, Trost, and Newsom (1995), and two from the CAT-Personality Disorder Scales Static Form (Wright, & Simms, 2014) ("*… pays attention to detail", "*… reflects on things before acting"). Reliability concerns how predictable and stable a person is in their actions. Items such as *" *... reflects on things before acting"* were used to assess reliability. This scale was scored by adding all the ratings together and dividing the total by seven. The present study found a coefficient alpha of .90 for this new scale.

Testing Procedure

Participants were given a page of information about the study and a consent form (See Appendix A). Once their signed consent had been obtained, participants were instructed to complete the SBT on the computer in front of them. Each participant was told to "*imagine you have been given this test as part of applying for a job. Your performance on this test will impact whether you are selected or not, so try your best."* After the participant completed the SBT they were given a usability questionnaire to complete. Once the usability questionnaire was finished the participant was given a $10 MTA voucher to thank them, and an acquaintance questionnaire to give to an acquaintance. The acquaintance sheet had instructions telling the acquaintance about the study, and what was required of them (see Appendix B). Individuals who returned Acquaintance questionnaires were given another $10 MTA voucher to give to the acquaintance for their time.

Study 2 Results

Data Management

The SBT data from each participant, data from their useability questionnaire, and data from their acquaintance questionnaire were input into an SPSS file containing, data from the Burt, Crowe, and Thomas (2018) validation study. Planned analysis called for a case to be removed if more than 10% of the data was missing, or if any of its values were more than three standard deviations away from the mean. Inspection of the data revealed some missing values, though no outlier cases. Two cases were removed from analysis for missing data, as the SBT was unfinished. Missing values were replaced with the item mean, so long as the missing values were not demographic data. Table 10 presents a summary of missing items, how many of each were missing, and the item mean used as a replacement value.

Table 10.

Missing Responses from study 2

Missing responses for each Measurement item	*Missing item*	*Number*	*Replacement Mean*
Acquaintance Questionnaire			
Safety Consciousness	Item 5	1	2.76
	Item 7	1	4.05
Risk Taking	Item 1	1	2.32
	Item 4	1	2.23
	Item 9	1	2.58
Rule Bending	Item 3	1	2.47
Reliability	Item 1	1	4.02

Distribution and range restriction in the SBT score

The SBT score data from studies 1 and 2 (N = 95) was examined in order to see whether the distribution, range, kurtosis and skew was of a level appropriate for correlational analyses. See table

11 for SBT score means and range. Plotting Study 1 and 2's SBT scores by frequency reveals a normally distributed bell curve (see Table 12, and Figures 13 and 14). The majority of participants performed close to the mean, with fewer participants at each end. Plotting Study 2 reveals some range restrictions, with no participants scoring lower than 4.

Table 11.

SBT Score Mean and Range for present studies

	SBT score Mean (SD)	Range
Study 1 & 2	8.32 (2.59)	0-13
Study 2	8.86 (2.18)	4-13

Table 12.

Frequency of each SBT Score

	Frequency of score	
SBT Score	*Study 1 and 2 n=95*	*Study 2 data n=35*
0	1	0
1	0	0
2	0	0
3	1	0
4	7	1
5	8	1
6	4	4
7	12	3
8	12	5
9	19	9
10	8	2
11	15	6
12	5	3
13	3	1

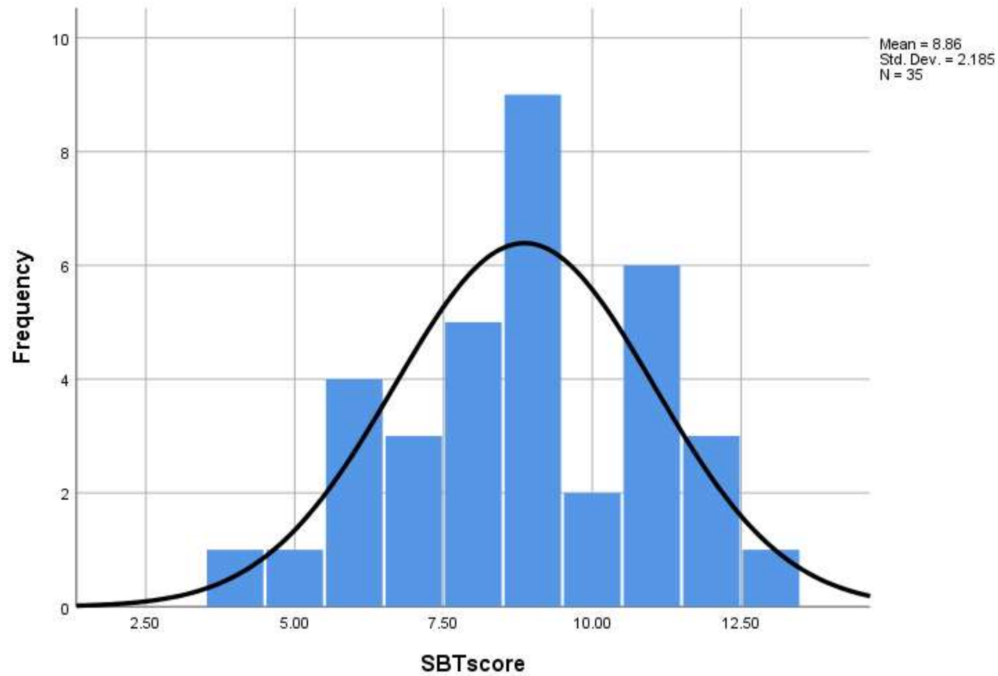


Figure 13. *Histogram of Study 2 SBT scores used in validation.*

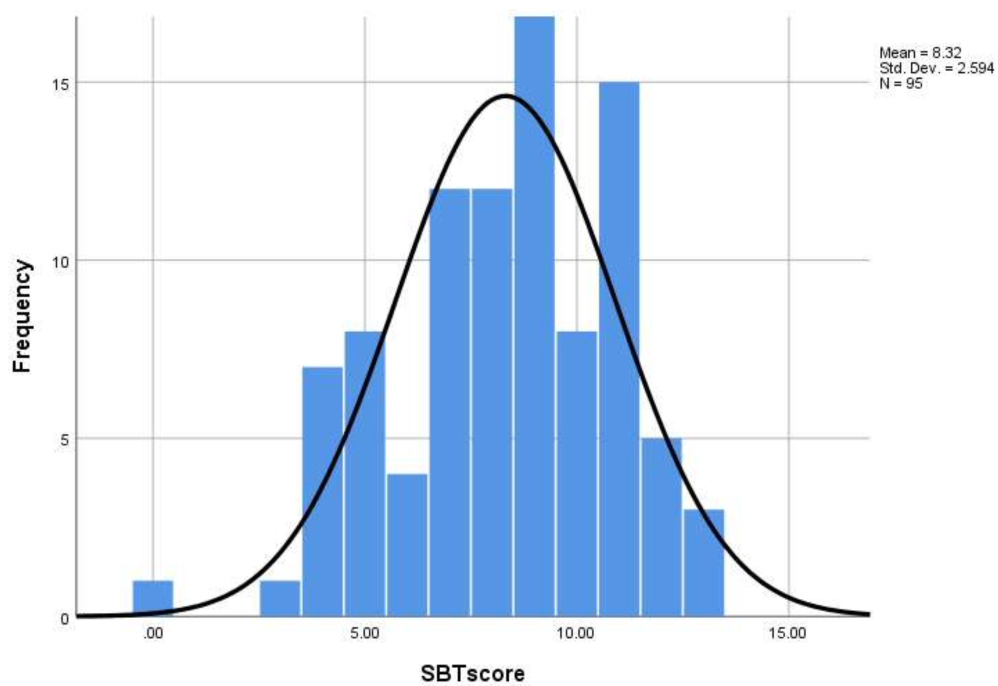


Figure 14. *Histogram of complete SBT 2.0 scores*

Statistical analysis confirms that Study 2 data shows some evidence of range restriction, and also some moderate negative kurtosis. Table 13 presents SBT score skew, kurtosis, and range restrictions for Study 2 (the validation study), study 1 & 2 data (all the data using SBT version 2.0), and Burt, Crowe, and Thomas (2018). All of the SBT 2.0 data (combined studies 1 and 2) shows normal amounts of skew and kurtosis, and no range restrictions. These results are consistent with a normal distribution. A normal distribution is encouraging for the newer version of the SBT's use as a measure. It is possible that Study 2 data suffered from range restriction due to a smaller sample size (Sackett, & Wade, 1983).

Table 13.
Distribution and Range Restriction in the SBT Score for Present Research, Study 2 and Burt, Crowe, and Thomas (2018)

			Range				
SBT Version	*n*	M (SD)	Possible	Actual	Skew	Kurtosis	Percentile SBT score (% below)
Burt, Crowe, and Thomas (2018)	100	8.05 (2.72)	0-13	0-13	-.45	-.06	6 (25%) 8 (50%) 10 (75%)
Study 2 Data	35	8.85 (2.18)	0-13	4-13	-.18	-.51	7 (25%) 9 (50%) 11 (75%)
Study 1 & 2 Data	95	8.31 (2.59)	0-13	0-13	-.47	-.00	7 (25%) 9 (50%) 10 (75%)

SBT participant and Acquaintance relationship

Before examining the criterion variables it was important to establish that the data source was valid in terms of knowing the participant for a sufficient amount of time and in a context which provided for valid rating of their safety behaviour. The relationships between SBT participants and their acquaintance were examined in terms of strength and type, see Table 14. In order for the acquaintance safety data to be representative of the participant's safety behaviours, the acquaintance would have to know the participant well. Table 14 reveals that the average relationship strength was quite high. Table 15 shows that the majority of relationships were work-based, and/or friendships.

This is promising, because work-based acquaintances would be better able to comment on the participants work safety behaviours. The strength of the relationships appears to be generally high also, which is promising.

Table 14.

Mean Acquaintance Relationship Length and Strength

Acquaintance Variables	*n*	M (SD)	Minimum	Maximum
Months known	35	67.6 (105.86)	2	480
How well known (0 - 100)	35	70.28 (21.45)	20	100

Table 15.

Frequency of Different Acquaintance Relationship Types

Acquaintance Relationship Type	Frequency
Work Colleague	25
Work Manager/Supervisor	3
Partner	4
Parent	1
Friend	14

Distribution and Range restriction in the criterion Variables

Safety behaviour criterion data was also examined for skew, kurtosis, and range restrictions (see Table 16). Safety behaviour criterion data was obtained with the following six scales from the acquaintance questionnaire: Reliability, safety consciousness, safety participation, safety compliance, risk-taking, and rule-bending.

Table 16.

Means, Distribution, and Range Restriction analysis for the Criterion Variables

Criterion Variables	*N*	Mean	Range	Skew	Kurtosis
Reliability	35	4.31	2.29 - 5.00	-1.62	3.92
Safety consciousness	35	3.89	2.29 - 4.86	-.83	2.50
Safety Participation	35	3.75	2.00 - 5.00	-.12	.93
Safety Compliance	35	4.16	2.67 - 5.00	-.33	.14
Risk-taking	35	2.43	1.00 - 4.40	.55	.21
Rule-Bending	35	2.57	1.00 - 4.25	.30	-.96

Inspection of Table 16 shows a large negative skew for the reliability measure, and a moderate negative skew for the safety consciousness measure. Both the reliability and safety consciousness scores also had large positive kurtosis. These factors mean that these measures tend away from normal distribution, and will likely supress correlations between these criterion variables and SBT scores (Field, 2009; Trochim & Donnelly, 2006;)

Criterion variable Relationships

The variables used in the acquaintance questionnaire to assess each participant's level of overall safety were chosen because together they encompass all aspects of safety behaviour. Reliability, Safety consciousness, Safety participation, safety compliance, risk-taking, and rule-bending all together provide the criteria against which the validity of the SBT is assessed. The correlations of these measures with each other is presented in Table 17. In addition, the relationship between the acquaintance rating of participant safety risk and the criterion variables was also examined.

Table 17.

Correlations between Safety Criterion Variables, and Acquaintance rating of risk.

Criterion Measures	1.	2.	3.	4.	5.	6.
1. Reliability	-					
2. Safety Consciousness	.45**	-				
3. Participation	.14	.57**	-			
4. Compliance	.31◦	.77**	.59**	-		
5. Risk-taking	-.09	-.58**	-.31◦	-.51**	-	
6. Rule-Bending	-.28	-.72**	-.35*	-.64**	.74**	-
7. Acquaintance rating of risk	-.22	-.51**	-.37*	-.43**	.43*	.58**

◦ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Inspection of Table 17 indicates that the reliability measure had the fewest number of significant correlations with other criterion variables, only correlating significantly with safety consciousness. This suggests that reliability captures a unique aspect of behaviour, meaning that its inclusion helps tap into behaviours not captured by the other variables. Reliability was also the only criterion variable not to correlate significantly with acquaintance rating of risk.

Safe behaviours (safety participation, safety compliance, and safety participation) all had significant positive correlations with each other, and significant (or large) negative correlations with rule-bending and risk-taking. Unsafe behaviours (risk-taking, and rule-bending) were significantly positively correlated. These correlations are as expected for the measured constructs. The acquaintance's rating of the participant's safety risk, a 0 to 100 rating of the participants "general degree of safety risk", appeared to be an excellent measure of participant safety behaviours. Inspection of Table 17 shows that the acquaintance rating of risk has significant negative correlations with the positive safe behaviours, and significant positive correlations with the unsafe behaviours.

Overall, the correlations between the variables are of a size consistent with the assumption they are each measuring unique aspects of the same construct: safety behaviour.

Criterion-Related Validation Analysis of the SBT Version 2.0

The SBT was validated by examining the correlations between participant SBT scores and of the six safety behaviour criterion measures reported by the acquaintance. Table 18 presents the correlations between SBT score, the six criterion variables, and acquaintance rating of risk. There were no significant correlations between any of the criterion measures and the SBT score. This is true even once video game experience is controlled for using a partial correlation.

Table 18.

Correlations between Safety Criterion Measures and SBT Scores Before and After Controlling for Videogame Experience

	SBT Score	
Criterion measure	All participants *n = 35*	Controlling for months playing videogames *n = 15*
Reliability	.23	.24
Safety consciousness	.13	-.03
Participation	-.04	-.20
Compliance	.09	-.08
Risk-taking	.15	.36
Rule-bending	-.11	-.43
Acquaintance rating of risk	-.00	.15

The results in Table 16 suggest that the SBT may have lost validity as a measure after the introduction of the tutorial. Previously Burt, Crowe, and Thomas (2018) had found that once video game experience had been controlled for, the SBT had validity as a measure of safety behaviours. Table 17 shows a comparison of validation analysis between the present study and Burt, Crowe, and Thomas (2018). Note the reliability measure was not used in the Burt, Crowe, and Thomas (2018) study, and voicing measuring was not used in the present study.

Table 19.

Correlation between SBT Scores and Criterion related Variables Before and After Controlling for Videogame experience across SBT Versions

	SBT Score			
	Present Study		Burt, Crowe, and Thomas (2018).	
Criterion measure	All participants *n = 35*	Controlling for months playing videogames *n = 15*	All participants *n = 100*	Controlling for months playing videogames *n = 30*
Reliability	.23	.24	-	-
Safety consciousness	.13	-.03	.08	.42*
Participation	-.04	-.20	.00	.22
Compliance	.09	-.08	.13	.42*
Risk-taking	.15	.36	-.08	-.41*
Rule-bending	-.11	-.43	-.20*	-.46**
Safety Voicing	-	-	.02	.25

* $p < 0.05$, ** $p < 0.01$

What happened?

The results may appear as though they suggest that there is no longer an adverse impact caused by a lack of computer game experience. However, the small sample size may be supressing correlations. The relationship between SBT scores and criterion variables still change when computer game playing experience is controlled for, but these correlations are not significant with a sample size of 15 as in the present study.

To look more closely at the impact of the new tutorial the participants user reaction ratings where correlated with the SBT scores. The samples using SBT version 1.0 were compared against those using SBT 2.0. Inspection of Table 18 shows that a factor that correlates with SBT scores is how *understandable* participants found the instructions. Understanding the instructions appeared to play a role in performance in the new version of the SBT, but not in Burt, Crowe, and Thomas' version (2018). Table 18 shows the relationship between participant feedback and performance between studies, for all participants, and after controlling for videogame experience. Table 18 suggests that the tutorial may have helped to increase performance, but only if the participant was able to understand it.

Table 18. *Comparison between Correlations of SBT Score with Participant Feedback across SBT Versions*

	SBT Score			
	Study 1 and study 2.		Burt, Crowe, and Thomas (2018)	
Feedback Measure	All SBT 2.0 Participants (Study 1 & 2) *n=95*	Controlling for months playing videogames *n=56*	All SBT 1.0 Participants (Study 1) *n=100*	Controlling for months playing videogames *n=30*
Understand Instructions	.27**	.38*	.10	.11
Easy to control forklift	.01	.08	.13	.05
Forklift appropriate speed	.23*	.19	.15	.00
Overall how easy	.05	.10	.06	.06
Enjoyment	.17	.21	.02	.09

* *p < 0 .05, ** p < 0.01*

Discussion

Summary

Previous research using the SBT found that video game experience had an adverse impact on the criterion related validity of the SBT. The assessment only obtained a criterion valid measure of behaviour from individuals who had video game experience. The present study introduced an interactive tutorial in an attempt to remove the adverse impact of video game experience on the SBT.

The study results showed that the tutorial removed the significant difference in SBT scores between those with computer game experience, and those without. However, while this remained true in the validation study (Study 2), SBT 2.0 scores did not correlate significantly with any of the safety criterion variables. Furthermore, non-significant correlations remained even once video game experience was controlled for.

The discussion will first examine the results associated with the introduction of the new pre-assessment tutorial. Then issues of criterion related validity will be examined, including limitations associated with the sample size, and the independent criterion measures of the participants' safety behaviour. The applied value of using an objective gamified measure of safety behaviour within the context of employee selection is then discussed. Finally, future research directions are noted.

The Pre-assessment SBT Tutorial

Interestingly there are variables that correlate with SBT scores in version 2.0, but not version 1.0. Participants rating of how *understandable* they found the pre-assessment instructions correlates with their SBT score in version 2.0. The more understandable the individual found the instructions, the higher their SBT score ($r = .27, p < .01$). When controlling for computer game experience the relationship between understanding instructions and SBT score becomes stronger ($r = .38, p < .05$). In other words, the new interactive pre-assessment tutorial appears to have helped individuals to increase their SBT score.

Perhaps the effect from *understanding* arises due to the comprehensive nature of the new pre-assessment interactive tutorial. The tutorial in SBT version 2.0 is exhaustive, and spells out exactly what participants are required to do to control the assessment. A participant can practise every aspect of assessment control, as many times as they like. The comprehensive tutorial may have mean that participants who understood how to control the assessment could have repeated their actions from the tutorial, and therefore had fewer issues with *control* decisions, instead focusing their mental efforts on *safety* decisions.

Ratings of understanding of the static tutorial sheet in SBT version 1.0 did not correlate with SBT score. Indeed, the instruction sheet in version 1.0 may have been unhelpful enough that only those with previous video game experience had sufficient understanding about how to interact with the SBT. In version 1.0 those without previous game experience may have had to spend their time in the assessment attempting to figure out how to interact with it, rather than acting authentically. This might explain why only those with video game experience were able to exhibit authentic behaviour, the way in which those with video game experience interacted with version 1.0 supports this theory. Individuals who tended to bend rules in their life, also bent the rules slightly in the assessment, making many more clicks over the limit rather than attempting to play the assessment using the click limit as instructed. In those with video game experience, the number of clicks they made correlated with their acquaintance-rated rule bending score ($r = .40, p < .05$). Individuals considered more thoughtful, and safety conscious by their acquaintance took more time to complete the assessment, presumably because they were considering their actions more. Time taken to complete the SBT 1.0 correlated with game player's safety consciousness score ($r = .34, p < .05$).

In context

The ability of the new pre-assessment tutorial to remove the difference between game player and non-game player SBT scores provides evidence against a hypothesis raised by earlier SBT research. Burt, Crowe, and Thomas (2018) suggested the possibility that the SBT score advantage associated with computer game players could be due to characteristics they possess, rather than their game control familiarity. While individuals with videogame experience have been found to possess higher levels of certain traits (for example: cognitive flexibility (Grottfredson, 2004), and intelligence (Pillay, 2002)), the present studies' findings that there are no significant difference in SBT scores between those with and without computer game experience appears to preclude this theory.

Version 2.0: Criterion Related Validity

As an assessment of safety behaviour, SBT Version 2.0, as assessed in Study 2, shows limited evidence of criterion related validity. However, the fact that in Version 2.0 there is no significant difference in scores between video game players and non-video game players mean that with the new pre-assessment tutorial it is possible to remove the advantage that game players have over non-game players. The advantage of those with video game experience in the field of gamified assessment is a known issue. Similar to Burt, Crowe, and Thomas (2018) with version 1.0 of the SBT, Kim and Shute (2015) also found that individuals who identified as gamers had an advantage on a gamified physics assessment. That Version 2.0 of the SBT was able to remove this advantage through a tutorial is a significant finding for the field of gamified assessments. This means that experience with computer games may not be an obstructing factor in the future application of the SBT for selection.

Issues

The key issue with the present findings is the reason as to why the SBT 2.0 failed to find criterion-related validity evidence. That the SBT version 2.0 is an improvement over version 1.0, and yet did not improve as an assessment, suggests that a closer examination of the independent criterion variables is warranted. The ability to show criterion-related validity rests heavily on the validity of the criterion data.

Criterion-Related Validity

The validation study found no significant correlation between any of the criterion related variables, and SBT score. While not significant, the correlations between SBT score and certain safety criterion variables in Study 2 do increase in size once computer game experience is controlled for. However, with the exception of the rule-bending score, these larger correlations are the opposite direction of the correlations found during study 1. Study 2 SBT score correlates positively with risk-taking, and negatively with safety participation, the complete opposite of the findings from study 1. These results are inconclusive, and probably subject to issues of sample size and sampling.

Skew and Kurtosis

During examination of the data evidence of skew was found for three of the acquaintance criterion variables. Large negative skew for the reliability measure, a moderate negative skew for the safety consciousness measure, and a low positive skew for the risk-taking measure were discovered. In addition, the reliability and safety consciousness scores also had large positive kurtosis. Together these factors represent data trends that will likely suppress correlations between these criterion variables and SBT scores (Trochim & Donnelly, 2006; Field, 2009)

The skew suggests that there may have been bias in the responding of the acquaintances. As mentioned earlier self-report questionnaires are prone to individual bias in an attempt to present an image of oneself. The skew present in the acquaintance data suggests that acquaintance questionnaires are subject to similar bias too. Perhaps the acquaintances had some bias when reporting on the behaviours of their colleagues, and friends. To say that another individual that they spend time with consistently acts unsafely may negatively impact the impression they have of themselves. Using an objective measure of safety, such as number of accidents and near misses over time would be a more accurate measure a participant's safety behaviour.

The present research was limited in terms of sample size, and comparisons between SBT Version 1.0 and 2.0 suffered for this. Where the study by Burt, Crowe, and Thomas (2018) had 100 participants, each with acquaintance criterion scores, Study 1, and 2 had 95 participants, with only 35 cases complete with acquaintance criterion scores in Study 2. This was caused, in part, because Study 1 and 2 had a smaller budget, and fewer researchers. However, practical concerns do not change the fact that the low sample size harms comparisons between the groups. The limited sample size also prohibits the correlations from being significant. In study 2, once computer game experience was controlled for, correlations as high as .43 were reached. However, once computer game experience is controlled for only 15 participants remain for analysis. A correlation of .43 could only be found significant at $p < .05$ level with a minimum sample size of 19 participants. To reach the more robust *p*-value of .01, the sample size would have to have been at least 33. The sample size in the validation study was therefore much too small when the typical correlations for assessment tools are taken into consideration. It is not possible to say whether the correlations would have stayed as large with the addition of more participants, though more participants are needed for future studies in order to prevent this issue.

Future safety behaviour assessment

Despite the SBT version 2.0 scores not showing significant correlations with the criterion data gamified assessment of safety behaviour holds far more promise as a selection tool than self-report tools, for the reasons outlined earlier. Individuals are prone to present an idealised version of themselves in a self-report survey. This positive bias is increased when there is an obvious desirable answer, and an advantage gained for responding in a specific manner (Leary, & Kowalski, 1990; Villanova, & Bernardin, 1991). This is incredibly pertinent to safety behaviour assessment tools in a selection context, where it is obvious that the correct answer is to present oneself as safe, and doing so will advantage the individual in obtaining the job. Additionally, self-reports are also only able to measure safety adjacent traits, and never the safety behaviours themselves. These safety adjacent traits such as safety knowledge, and personality only correlate weakly with accidents in the workplace (Christian et al., 2009)

For these reasons, and despite the weak findings in the present research, it is imperative that the SBT continues to be developed, and validated in a better fashion. A quality gamified assessment would completely avoid the pitfalls of self-report safety behaviour assessments.

Future SBT research

Clearly, there is a need for a predictive validity study of the SBT version 2.0. This study would greatly increase the criterion side of the validation work by examining the relationship between SBT score and objective safety outcomes such as, a measure of workplace accidents. Such an approach would mean that the criterion measures would have reduced bias. In addition, it would be a good way of directly assessing the SBT as a selection tool. The desired outcome is to use the SBT to assess an individual’s likelihood to have a workplace accident. A predictive validation study would replicate the SBT’s future application much more closely than the present concurrent validation research.

Future validation studies of the SBT 2.0 would need to obtain more participants, taking care to ensure that there is a suitable number of individuals with computer game experience, and a similarly sized group without. A minimum of 35 individuals in each computer experience group would ensure that the correlations typically found in such studies of ~.40 could be significant at the $p < .01$ level. A larger sample size would also reduce the effect that individual variance could have on the results.

Conclusion

There is still some suggestion in the validation study that controlling for computer game experience affects the authentic nature of the relationship between SBT score and safety behaviour. No definitive results were found, due to sampling size, and skew in the data. However it appears that the pre-assessment tutorial was helpful for participants. Compared to the SBT 1.0, there were no significant differences between mean SBT scores of those with and those without computer game experience. Understanding the tutorial appeared to allow individuals more ability to control the assessment, and make better safety decisions. Further research should examine the SBTs predictive validity, and seek larger sample sizes.

References

Accident Compensation Corporation. (2018). Statistics. Retrieved from http://www.acc.co.nz/about-acc/statistics/

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal*, *24*(2), 377-385.

Bailey, P., Pritchard, G., & Kernohan, H. (2015). Gamification in market research: increasing enjoyment, participant engagement and richness of data, but what of data validity?. *International Journal of Market Research*, *57*(1), 17-28.

Barling, J., Loughlin, C., & Kelloway, E. K. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of applied psychology*, *87*(3), 488.

Briggs, P. (2003). Looking at the numbers. *A view of New Zealand economic history. Wellington: NZIER*.

Barrett, P. (2010). A review of commercial products and academic articles associated with the psychological assessment of Safety Attributes within prospective and incumbent employees. Auckland, New Zealand: Advanced Projects R&D Ltd. Retrieved October, 19, 2018, from http://www.pbarrett.net/Safety_Review/A_Review_of_the_Psychological_Assessment_of_Safety_Attributes_Paul%20Barrett_Oct_2010.pdf

Burt, C. D. B., (2015). New Employee Safety: Risk Factors and Management Strategies. Springer.

Burt, C. B. D., Crowe, L., & Thomas, K. (2018). Validation of a gamified measure of safety behavior: The SBT. In *Safety and Reliability–Safe Societies in a Changing World* (pp. 263-270). CRC Press.

Chmiel, N. (2005). Promoting health work: Self-reported minor injuries, work characteristics, and safety behaviour. In C. Krunka, P. Hoffmann, & A Bussing (Eds.), Change and quality in human service work (p 277 – 288). Mering: Rainer Hampp Verlag. Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: a meta-analysis of the roles of person and situation factors. Journal of Applied Psychology, 94(5), 1103.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioural implications. *Journal of personality and social psychology*, *69*(2), 318.

Clarke, J. (2009). Studying the potential of virtual performance assessments for measuring student achievement in science. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Collmus, A. B., & Landers, R. N. (2015). Game narrative in personality assessment: The development of a scale. In *Virginia Psychological Association's Spring Convention and Education Conference, Virginia Beach, VA*.

DeFreitas, G., & Marshall, A. (1998). Labour surplus, worker rights and productivity growth: a comparative analysis of Asia and Latin America. *Labour*, *12*(3), 515-539.

Department of Labour (2003). A Guide to the Health and Safety in Employment Act 1992 Including the Health and Safety in Employment Amendment Act 2002. Retrieved from:https://cdn.auckland.ac.nz/assets/science/for/current-students/HR/health-safety wellness/documents/hseguide-2ed.pdf

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011, May). Gamification, using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425-2428). ACM.

Easton, B. (2016) Economy - Agricultural production. Retrieved from:http://www.TeAra.govt.nz/en/diagram/42316/percentage-of employment-in different-sectors-1841-2001 (accessed 12 August 2018)

Embrick, D. G., Wright, T. J., & Lukács, A. (Eds.). (2012). *Social exclusion, power, and video game play: New research in digital media and technology*. Lexington Books.

Evans, L., & Quigley, N. (2003). Accident Compensation: The Role of Incentives, Consumer Choice and Competition. *Victoria U. Wellington L. Rev.*, *34*, 423.

Feldman, D. C., & Arnold, H. J. (1978). Position choice: Comparing the importance of organizational and job factors. *Journal of Applied Psychology*, *63*(6), 706. Field, A. (2009). Discovering statistics using SPSS. Sage publications.

Gill, G. K., & Shergill, G. S. (2004). Perceptions of safety management and safety culture in the aviation industry in New Zealand. *Journal of Air Transport Management*, *10*(4), 231-237.

Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive" fundamental cause" of social class inequalities in health?. *Journal of personality and social psychology*, *86*(1), 174.

Hafer, C. L., Reynolds, K. L., & Obertynski, M. A. (1996). Message comprehensibility and persuasion: Effects of complex language in counterattitudinal appeals to laypeople. *Social Cognition*, *14*(4), 317-337.

Hämäläinen, P., Takala, J., & Kiat, T. B. (2017). Global estimates of occupational accidents and work-related illnesses 2017. *World*, *2017*, 3-4.

Heinrich, H. W. (1931). Industrial accident prevention: A scientific approach. New York: McGraw-Hill.

Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., & Robinson, E. V. (2005). Construct validity of the balloon analog risk task (BART) associations with psychopathy and impulsivity. *Assessment*, *12*(4), 416-428.

International Labour Organisation. (2017). Safety and Health at Work. Retrieved from http://www.ilo.org/global/topics/safety-and-health-at-work/lang--de/index.htm

Kim, H. T., Na, S., & Ha, W. H. (2011). A case study of marine accident investigation and analysis with focus on human error. *Journal of the Ergonomics Society of Korea*, *30*(1), 137 -50.

Kim, Y. J., & Shute, V. J. (2015). Opportunities and challenges in assessing and supporting creativity in video games. In *Video games and creativity* (pp. 99-117).

Kocadere, S. A., & Çağlar, Ş. (2015). The design and implementation of a gamified assessment. *Journal of e-Learning and Knowledge Society*, *11*(3).

de Koster, R. B., Stam, D., & Balk, B. M. (2011). Accidents happen: The influence of safety specific transformational leadership, safety consciousness, and hazard reducing systems on warehouse accidents. *Journal of Operations management*, *29*(7-8), 753 765.

Kozlov, M. D., & Johansen, M. K. (2010). Real behavior in virtual environments: Psychology experiments in a simple virtual-reality paradigm using video games. *Cyberpsychology, behavior, and social networking*, *13*(6), 711-714.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*(1), 34.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of adolescence*, *26*(4), 475-479.

Lejuez, C W., Read, J., Kahler, C., Richards, J., Ramsey, S., Stuart, G. L., Strong, D., & Brown, R. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). Journal of experimental psychology. Applied. 8. 75-84. 10.1037/1076-898X.8.2.75.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety science*, *42*(4), 237-70.

Lilley, R., Samaranayaka, A., & Weiss, H. (2013). International comparison of International Labour Organisation published occupational fatal injury rates: How does New Zealand compare internationally. *Commissioned report for the Independent Taskforce on Workplace Health and Safety*.

MacEachen, E., Kosny, A., Ferrier, S., & Chambers, L. (2010). The "toxic dose" of system problems: why some injured workers don't return to work as expected. *Journal of occupational rehabilitation*, *20*(3), 349-366.

McNoe, B., Langley, J., Driscoll, T., & Feyer, A. M. (2005). *Work-related slip, trip and fall injuries in New Zealand*. Injury Prevention Research Unit.

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40.

Neal, A., & Griffin, M. A. (2002). Safety climate and safety behaviour. *Australian journal of management*, *27*(1_suppl), 67-75.

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of applied psychology*, *91*(4), 946.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European journal of social psychology*, *15*(3), 263-280.

NZ Statistics. (2018). I*njury Statistics, Work related claims.*Retrieved from:https://www.stats.govt.nz/insights?filters=Injury%20statistics%20%E2%80%9 %20work-related%20claims%2CInformation%20releases

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology*, *46*(3), 598.

Pezzullo, L., & Crook, A. (2006). *The economic and social costs of occupational disease and injury in New Zealand*. NOHSAC.

Pillay, H. (2002). An investigation of cognitive processes engaged in by recreational computer game players: Implications for skills of the future. *Journal of research on technology in education*, *34*(3), 336-350.

Pinelle, D., Wong, N., & Stach, T. (2008, April). Heuristic evaluation for games: usability principles for video game design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1453-1462). ACM.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213.

Reason, J. (1990). *Human error*. Cambridge university press.

Sackett, P. R., & Wade, B. E. (1983). On the feasibility of criterion-related validity: The effects of range restriction assumptions on needed sample size. *Journal of Applied Psychology*, *68*(3), 374.

Salminen, S., Klen, T., & Ojanen, K. (1999). Risk taking and accident frequency among Finnish forestry workers. *Safety Science*, *33*(3), 143-153.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, *82*(5), 719.

Shappell, S. A., & Wiegmann, D. A. (1996). US naval aviation mishaps, 1977-92: differences between single-and dual-piloted aircraft. *Aviation, Space, and Environmental Medicine*, *67*(1), 65-69.

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. W. (2016). Advances in the science of assessment. *Educational Assessment*, *21*(1), 34-59.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, *106*(6), 423-430.

Slappendel, C., Laird, I., Kawachi, I., Marshall, S., & Cryer, C. (1993). Factors affecting work-related injury among forestry workers: A review. *Journal of safety research*, *24*(1), 19-32.

Stanton, N. A., & Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, *47*(2), 227-237.

Suh, A., Wagner, C., & Liu, L. (2015, January). The effects of game dynamics on user engagement in gamified systems. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 672-681). IEEE.

Trochim, W., & Donnelly, J. (2006). The research knowledge methods base. *Cincinnati, OH: Atomic Dog Publishing*.

Turner, C., McClure, R., & Pirozzo, S. (2004). Injury and risk-taking behavior—a systematic review. *Accident Analysis & Prevention*, *36*(1), 93-101.

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), 40.

Villanova, P., & Bernardin, H. J. (1991). Performance appraisal: The means, motive, and opportunity to manage impressions.

Wagner, M. K. (2001). Behavioral characteristics related to substance abuse and risk-taking, sensation-seeking, anxiety sensitivity, and self reinforcement. *Addictive behaviors*, *26*(1), 115-120.

Westaby, J. D., & Lee, B. C. (2003). Antecedents of injury among youth in agricultural settings: A longitudinal examination of safety consciousness, dangerous risk taking, and safety knowledge. *Journal of safety research*, *34*(3), 227-240.

Wright, A. G., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analyses of the CAT-PD, PID-5, and NEO-PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment*, *5*(1), 43.

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: common biosocial factors. *Journal of personality*, *68*(6), 999-1029.

Appendix A
Usability Questionnaire

Department of Psychology
Email: chrisd.watson@pg.canterbury.ac.nz

31/05/2018

Validation of the SBT: The Impact of Individual Characteristics on SBT SBT Participant Information Sheet

I am Chris Watson, and I am a Masters student in the Department of Psychology at the University of Canterbury. The purpose of this research is to validate the *Safety Behaviour Test (SBT)*. The SBT is a fully animated computer game. The players must point the cursor at areas on the screen and click in order to interact with the game environment. In the SBT, players will be given instructions to retrieve several different items from within a warehouse using a forklift, and then load each item into a container. In order to validate the SBT, the current study will require participants to complete both the SBT, and a questionnaire. The results of the questionnaire will be used to determine if any individual characteristics have an identifiable impact on SBT use and performance.

If you choose to take part in this study, your involvement in this project will be to complete the SBT, and to complete a questionnaire that assesses individual characteristics. The SBT and the questionnaire will each take approximately 20 minutes to complete.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you. However, once analysis of raw data starts on the 1st of November *2018*, it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. Furthermore, all physical data will be stored in a locked filing cabinet in a locked room, while all electronic data will be stored in a password protected computer in a locked room, and no person outside of the research team will have access to data. A thesis is a public document and will be available through the UC Library. Data will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out in partial fulfilment of the requirements for the degree of Master of Science in Applied Psychology at the University of Canterbury by Chris Watson under the supervision of Associate Professor Christopher Burt, who can be contacted at christopher.burt@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch(human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the researcher.

Department of Psychology
Email: chrisd.watson@pg.canterbury.ac.nz

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

31/05/2018

Validation of The SBT: The Impact of Individual Characteristics on SBT Use.

SBT Participant Consent Form

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and supervisor of the research and that any published or reported results will not identify the participants or organisation. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

☐ I understand that there are no risks associated with taking part in this study

☐ I understand that I can contact the researcher, Chris Watson (chrisd.watson@pg.canterbury.ac.nz) or supervisor Christopher Burt (christopher.burt@canterbury.ac.nz) for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800,Christchurch (human-ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:____________________________Signed:__________________________Date:________________

Email address *(for report of findings, if applicable)*:____

Test code ……………

1. **Demographic Questions**

- Your Age:………. - Gender:……….

2. **Computer Game Experience**

- Have you played a computer game before using the SBT?

□ Yes □ No

- Please indicate how many years and months you have been playing computer games.

  ……….Years and……….months

- Have you ever played a point and click game?

  ☐ Yes ☐ No ☐ Don't know

- How often do you play computer games?

  ☐ Daily ☐ Once every 6 months
  ☐ Weekly ☐ Once a year
  ☐ Monthly ☐ Less than once a year

3. **Work Experience**

- Do you work full time or part time? Full time □ Part time □ Don't work □ (go to section 4)

- Please indicate how many years and months you have had your current job for.

  ……….Years and……….months

- How many co-workers do you currently have?..........

- In total, how many different jobs have you had?..........

- Please indicate how many years and months you have worked for, in total.

  ……….Years and……….months

- How many different organisations have you worked for?..........

4. **SBT Usability:** The following questions are about your experience with the test you just completed.

- How understandable were the instructions given to you to use the SBT (please circle a number)?

  1……….2………..3………..4………..5……….6………..7

  Not at all Completely

- How easy was it to control the forklift in the SBT (please circle a number)?

  1……….2………..3………..4………..5……….6………..7

  Very Hard Very Easy

- How appropriate was the speed that the forklift moved in the SBT?

  1……….2………..3………..4………..5……….6………..7

  Very Inappropriate Very Appropriate

- Overall, how easy was it to complete the SBT?

  1……….2………..3………..4………..5……….6………..7

  Very Hard Very Easy

- How much did you enjoy completing the SBT?

  1……….2………..3………..4………..5……….6………..7

  Not at all Completely

- Do you have any other comments in regard to using the SBT (please write below)?

____

____

____

Please check that you have answered all questions.

Thank you for taking the time to participate in this research.

Appendix B
Acquaintance Questionnaire

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department of Applied Psychology
Email: chrisd.watson@pg.canterbury.ac.nz
Date: 1/11/18

Validation of the Safety Behaviour Test: Criteria Validity Evidence

Acquaintance Information Sheet

I am Chris Watson, a Masters of Applied Psychology student at the University of Canterbury conducting a study of the validity of the *Safety Behaviour Test (SBT)*. The purpose of the research is to establish if the SBT is a valid measure of safety compliance and participation.

The individual (*……………) who has given you this form has already had their safety behaviours assessed using a new computer assessment (the SBT). We would like to compare their safety behaviours in the test with their real world safety behaviours. If you choose to take part in this study, your involvement requires you to spend approximately 15 minutes completing the attached questionnaire. This questionnaire includes several safety behaviour items about *……………… who has consented to you completing this questionnaire about them. Whenever you see *… below this refers to the person who you are completing this questionnaire about. After completing this questionnaire and the consent form, return them to the researcher and collect your $10 petrol voucher.

Participation is voluntary and you have the right to withdraw at any stage without penalty. If you withdraw, I will remove information provided by you. However, once analysis of raw data starts, it will become increasingly difficult to remove the influence of your data on the results.
The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public. No one other than Chris Watson (as the researcher) and Chris Burt (as the research supervisor) will have access to the data. Physical data will be stored in a locked filing cabinet in a locked room. Electronic data will be stored on a password protected computer, in a locked room. Data will be destroyed after 5 years, unless a publication outlet requires extended archiving of the data. A thesis is a public document and the subsequent thesis will be available through the UC Library.

Please indicate on the consent form if you would like to receive a summary of results of the project. This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form before completing the questionnaire.

Department of Applied Psychology
Email: chrisd.watson@pg.canterbury.ac.nz
Date: 1/11/18

Validation of the Compliance and Participation Test: Criteria Validity Evidence

Acquaintance Consent Form

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

☐ I understand there are no risks associated with taking part in this study

☐ I understand that I can contact the researcher [*Chris Watson* chrisd.watson@pg.canterbury.ac.nz] or supervisor [*Chris Burt* christopher.burt@canterbury.ac.nz] for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:__________________________Signed:______________________________Date:____________

Email address *(for report of findings, if applicable)*:
____

Code: _______________

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department of Applied Psychology
Email: chrisd.watson@pg.canterbury.ac.nz
Date: 1/11/18

Questionnaire

Remember: *Whenever you see *... below this refers to the person who you are completing this questionnaire about*

General questions about you:

Your Age _____

Your Gender: _______________

How long have you known *… for? Years _____ Months_____

How do you know *… ? (tick as many categories as necessary)

I am *…'s
- Work colleague ☐
- Work manager ☐
- Worker supervisor ☐
- Spouse ☐
- Child ☐
- Parent ☐
- Friend ☐
- Sport/Recreation associate ☐
- Other ☐ (please specify_____________________________)

Please indicate how well you know *…. by placing a mark on the 100 point scale.

0.....10…..20…..30…..40…..50…..60…..70…..80…..90…..100

Not Very Well At All — Extremely Well

Considering *…'s behaviour in all the situations that you know, please indicate *…'s general degree of safety risk by placing a mark on the 100 point scale.

0.....10…..20…..30…..40…..50…..60…..70…..80…..90…..100

Not at all Risky — Extremely Risky

Safety Behaviour

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please circle a number to indicate how much you agree or disagree with each statement. If you don't know about any item please tick in the 'don't know' column.

	Strongly Disagree	Disagree	Neither Agree/ Disagree	Agree	Strongly Agree	Don't Know
*… always uses all the necessary safety equipment	1	2	3	4	5	
*… always uses the correct safety procedures	1	2	3	4	5	
*… always ensure the highest level of safety	1	2	3	4	5	
At work *…. promotes the safety programme within their organisation	1	2	3	4	5	
At work *… puts in extra effort to improve the safety of their workplace	1	2	3	4	5	
At work *… voluntarily carries out tasks or activities that help to improve workplace health and safety	1	2	3	4	5	

Safety Consciousness

These statements are about how *… behaves. For each statement, please circle the number which indicates the extent to which you disagree or agree. If you don't know about an item please tick in the 'don't know' column.

	Strongly Disagree	Disagree	Neither Agree/ Disagree	Agree	Strongly Agree	Don't Know
*… always take extra time to do things safely	1	2	3	4	5	
People think of *… as being an extremely safety-minded person	1	2	3	4	5	
*… always avoid dangerous situations	1	2	3	4	5	
*… takes a lot of time to do things safely even when it slows performance	1	2	3	4	5	
*… often makes sure that other people do things that are safe and healthy	1	2	3	4	5	
*… gets upset when seeing other people acting dangerously	1	2	3	4	5	
*… thinks doing the safest possible thing is always the best thing	1	2	3	4	5	
*… would rather take risks than be overly cautious	1	2	3	4	5	
In the past month *… has done some exciting things that other people might think are dangerous	1	2	3	4	5	
*… loves to take risks even when there is a small chance *… could get hurt	1	2	3	4	5	
Sometimes people get on *… nerves when they tell *… how to act "more safely"	1	2	3	4	5	
*… values having fun more than being safe	1	2	3	4	5	

Rule Breaking

These statements are about how *… behaves. For each statement, please circle the number which indicates the extent to which you disagree or agree. If you don't know about an item please tick in the 'don't know' column.

	Strongly Disagree	Disagree	Neither Agree/ Disagree	Agree	Strongly Agree	Don't Know
*… sometimes cuts corners if it makes the task easier	1	2	3	4	5	
Work pressures means the *… sometimes bends the rules	1	2	3	4	5	
Occasionally*….bends the rules when he/she knows it is safe to do so	1	2	3	4	5	
When *…'s boss is not around he/she can be more flexible with which procedures he/she follows	1	2	3	4	5	

Reliability

These statements are about how *… behaves. For each statement, please circle the number which indicates the extent to which you disagree or agree. If you don't know about an item please tick in the 'don't know' column.

	Strongly Disagree	Disagree	Neither Agree/ Disagree	Agree	Strongly Agree	Don't Know
*… is a reliable person	1	2	3	4	5	
*… pays attention to detail	1	2	3	4	5	
*… is a stable and predictable person	1	2	3	4	5	
*… reflects on things before acting	1	2	3	4	5	
*… can be counted on to get things done	1	2	3	4	5	
*… considers things carefully before acting	1	2	3	4	5	
*… makes an effort to be consistent	1	2	3	4	5	

Thank you for participating in this research, please return this form to the researcher.